7 Detection avoidance and enforcement theory
Chris William Sanchirico

1 Introduction
The subject of evidentiary foul play – inclusive of fabricated testimony, document destruction, and myriad other modes of detection avoidance – is underrepresented in both legal and law and economic scholarship on procedure and evidence. Judges and practitioners report that evidentiary misdeeds are commonplace (though systematic evidence is scarce). Explicit scholarly analysis of evidentiary misbehavior, however, is relatively uncommon.

This chapter considers attempts to add evidentiary misconduct to the conventional economic model of enforcement (Becker (1968) and others). In such context, evidentiary misconduct is typically referred to generically as “detection avoidance.” Another chapter in this volume, Evidence: Theoretical Models, reviews the chief formal approaches to legal evidence. To varying extents the models reviewed in that chapter also take on the issue of evidentiary foul play, though in a different manner.

This chapter begins in Section 2 with a brief overview of the rather complex legal landscape that covers evidentiary misconduct. This is followed in Section 3 with a description of how the economic framework conventionally employed in studying enforcement must be expanded to accommodate detection avoidance activities. Section 4 examines a particular consequence of this expansion in the case in which detection avoidance activities are not themselves sanctioned – namely, that maximal sanctions may no longer be optimal. Sections 5 and 6 consider the sanctioning of detection avoidance activities themselves. Section 6 also examines the difficult problems for a sanctioning approach that are caused by the recursive nature of detection avoidance. Section 7 considers the alternative “technological” approach to detection avoidance.

2 Brief Overview of Legal Landscape
Two kinds of legal rules regulate evidentiary conduct. The first consists of penalties for misconduct. The second consists of what might be called “technological” measures: roughly, policies that decrease the effectiveness of effort and expense devoted to detection avoidance.
2.1 Penalties for Misconduct

Penalties for evidentiary foul play are imposed by a complex and overlapping set of statutes, procedural rules, precedents, judicial practices, and professional codes.¹

First, three kinds of statutes criminalize evidentiary foul play: obstruction of justice, criminal contempt, and perjury. Criminal sanctions include fines and imprisonment. Second, when the underlying case is itself a criminal action, a convicted defendant who attempted to avoid conviction by means of evidentiary misconduct may face a stiffer sentence. US Federal Sentencing Guidelines, for example, provide for a sentencing enhancement for obstructing the investigation, prosecution, and sentencing of underlying offenses. Third, the court hearing the underlying case may impose sanctions for evidence tampering under either procedural rules (such as the US Federal Rules of Civil Procedure) or its “inherent power” to regulate process. Such sanctions include striking the offending party’s pleadings, entering a default judgment against the party, dismissing a party’s claims, forcing the party to pay her opponent’s attorneys’ fees, and providing a jury instruction in regard to what may be deduced from the fact that evidence has been destroyed. Fourth, a small number of jurisdictions allow private tort suits for evidentiary misconduct. Fifth, lawyers involved in evidentiary misconduct may be fined or disbarred under rules concerning professional responsibility.

Two regularities may be identified in the tangle of laws and rules penalizing evidentiary misconduct.² First, sanctions for evidentiary misconduct generally apply only to conduct that occurs far downstream in the litigation flow from primary activity through filing, discovery, and trial.³ For example, in the US, Federal Rule of Civil Procedure 37, which sanctions discovery misconduct, only comes into play after the complaint has been filed; perjury can only be committed under oath; and the so-called omnibus obstruction of justice provision, section 1503, requires that a proceeding be pending. Moreover, with regard to the exercise of the court’s “inherent power” to regulate its own proceedings – a term used to describe uncodified, though precedent-bound sources of judicial authority – a careful reading of the cases indicates that the courts have

---

¹ These are reviewed in detail in Gorelick et al. (1989), Koesel et al. (2000), and Sanchirico (2004a).
² Sanchirico (2004a) identifies this regularity in the context of Federal Civil Procedure in the US.
³ Sanchirico (2004a) suggests that this structure may well be justifiable given that the private efficiency of detection avoidance increases as litigation proceeds and the identity of decisive items of evidence comes into clearer view.
been reluctant to exercise such powers to sanction pre-filing destruction. This reluctance is all the greater when no specific plaintiff looms on the horizon. *A fortiori*, firms’ “document retention policies” that are not specifically directed at destroying potentially damaging records appear to remain largely effective as a means of insulating document destruction from inherent power sanctioning. Even section 1512(c), a new obstruction of justice provision, which makes general obstructive behavior criminal when there is no pending proceeding, may not extend to defendants who destroy evidence as part of routine document retention policies, or who, more generally, do not have a specific suit with a specific opponent in mind.

Second, even in the downstream reaches of the litigation flow wherein tampering is punished, the farther downstream the tampering, the more far reaching the prohibition. Of the two US federal perjury statutes, for example, section 1623, which prohibits lying only in judicial proceedings and depositions, is in other respects broader in scope than its counterpart section 1621, which applies any time a statement is made under oath. A similar pattern emerges from the untidy array of procedural and evidentiary sanctions. Here US Federal Rule of Civil Procedure 37(b) and inherent powers stand out as the main sources of sanctioning authority in practice. A court order to compel discovery under Rule 37 is only issued at the insistence of the opposing party. But if that order is later violated, penalties will be summarily and almost certainly imposed. If the court wishes to punish tampering somewhat farther upstream, it must use its inherent powers. The imposition of sanctions under this authority is hardly summary. The court must find that the offender had a “duty” to preserve the evidence, an inquiry which implicates the “reasonableness” of the destruction as well as the nexus between the destruction and the litigation. Moreover, as compared to the list of sanctions laid out in Rule 37(b), the typical inherent powers sanction is relatively lenient – an adverse inference instruction, which would seem to place the spoliator in the same position she would be in if she had not spoliated.

### 2.2 “Technological” Measures

In addition to penalizing evidentiary misconduct, certain aspects of investigation and evidentiary procedure may discourage such misconduct and/or render it less effective. These measures are helpfully categorized according to whether they exploit limits of human cognition or the limits of human cooperation. These two categories are discussed in more detail in Section 7.3.
3 How the Framework of the Enforcement Model Changes when Detection Avoidance is Incorporated

The probability that wrongs and offenses are detected is a central feature of the conventional economic model of enforcement. But the conventional model is starkly asymmetric regarding the determinants of this probability. It takes into account the government’s efforts at detecting violations, but it ignores violators’ efforts at avoiding detection. After briefly describing the conventional enforcement model, this section examines how incorporating detection avoidance activities changes the model.

3.1 The Conventional Approach (in Brief)

The conventional approach to enforcement – grounded in the work of Beccaria (1764), Bentham (1789) and Becker (1968) – has two essential characteristics: (1) its account of the basic “machinery” of deterrence, and (2) its description of the social cost-benefit analysis that ought to be conducted in making policy choices regarding public enforcement.4

3.1.1 Deterrence mechanics in the conventional approach Under the conventional approach to enforcement, the deterrent force exerted by law is viewed as the conjunction of two factors: the probability \( p \) that violations are “detected” and the magnitude of the sanction \( S \) imposed in the event of detection.7 The potential violator8 chooses to commit the violation if the expected sanction \( pS \) is less than her net private benefit \( b \) from committing the violation. The expected sanction \( pS \) may thus be regarded as the “level of deterrence.” In order to avoid unrealistic binary outcomes in which either everyone or no one commits the violation, potential violators are usually assumed to have heterogeneous private benefits. If the cumulative distribution function \( F(b) \) describes the population distribu-

---

4 The account provided here is necessarily cursory. For more complete descriptions of the conventional approach – and its many variants – see Garoupa (1997), Polinsky and Shavell (2007), and Franzoni (2000).
5 Unless otherwise noted, I shall be assuming throughout this discussion that violating the law is a binary choice. An important exception is the analysis in Section 4.2.
6 Unless otherwise noted, I shall be assuming that individuals who choose not to violate the law are never mistakenly sanctioned. A more general model produces results that are qualitatively similar to those discussed in this entry.
7 All variables that are not probabilities are expressed in terms of dollars. Concepts that are not natively measured in currency are implicitly converted to their dollar equivalents.
8 Unless otherwise noted, I shall be assuming that individuals are fully rational and risk neutral.
tion of the private benefits from violation $b$, then $1 - F(pS)$ is the fraction of the population of potential violators that actually violates the law. If $N$ is the size of the population of potential violators, then $N(1 - F(pS))$ is the number of violators.

3.1.2 Social cost-benefit analysis in the conventional approach

The basic machinery of deterrence described in Section 3.1.1 immediately above can be configured in many ways. For example, the size of the sanction, the nature of the sanction, and the frequency of detection are all subject to policy choice.

Under the conventional model, a particular cost-benefit analysis guides these policy choices.

The social benefits of deterrence are taken to be the benefits of reduced violations, including the benefits to those who would otherwise be victimized. Thus, the social benefit of establishing the level of deterrence $pS$ derives from the fact that some portion of potential violators is induced not to commit the violation. If $h$ is the level of aggregate net private harm imposed upon victims by each violation, and if it is assumed that the private benefits of violations are regarded as socially cognizable, then the net social benefit of the level of deterrence $pS$ equals $NF(pS)(h - E[b|b < pS])$, where $E[b|b < pS]$ is the (unrealized) average private benefit of violations among potential violators who choose not to violate the law. In this expression $NF(pS)$ is the number of violations that do not occur and $h - E[b|b < pS]$ is the average net social savings for each.

The social costs of deterrence are typically parsed into two categories, corresponding to the two factors in the conventional approach to deterrence mechanics, as described in Section 3.1.1 above. First, there are “detection costs,” the publicly incurred cost of investigating and prosecuting violations, as manifest in budgeting for regulatory enforcement divisions, police departments, and court systems. Thus, the probability of detection $p$ is regarded as an increasing function, $p(d)$, of such detection costs $d$. Second, there are “sanctioning costs,” $c(S)$, the cost of imposing sanctions when violations lead to conviction or liability, including, for example, the operating costs and opportunity costs of keeping convicts in prison. The cost of sanctions depends not just on the magnitude but also on the nature of the sanction. Prison time is costly, and more prison time is more costly. Fines are, however, generally regarded as costless transfers between individuals and the government; for fines $c(S)$ is constant at zero.

---

$^9$ Unless otherwise noted, it is assumed throughout that $h$ is uniform across violations.
3.2 *The Conventional Enforcement Model with Detection Avoidance Added*

3.2.1 The importance of adding detection avoidance

Although detection activities and their social cost play an important role in the conventional approach to enforcement, the conventional approach generally does not take into account detection *avoidance* activities and their costs. This asymmetry is difficult to justify. Although systematic statistical evidence is scarce, casual empiricism points to the conclusion that violators are more than mere spectators. Just as the state invests in detecting their violations, they invest in avoiding that detection. The investigation and prosecution of crimes and regulatory violations, it would seem, is not an exercise in orienteering, but a chase, consisting of a pursuit and a flight.

When detection avoidance activity is added to the conventional model of enforcement, both deterrence mechanics and social cost-benefit accounting are altered. I consider these effects in turn.

3.2.2 The effect of detection avoidance on deterrence mechanics

Under the conventional approach, the degree to which underlying violations are deterred depends on the sanction and the detection probability. Detection avoidance complicates this simple mechanic in four ways. These four ways may be grouped into two categories.

3.2.2.1 The effect on the level of deterrence for the underlying activity

3.2.2.1.1 First effect

First, and most obviously, detection avoidance activities reduce the probability that underlying violations will be detected. To this extent, such activities reduce the level of deterrence of underlying violations. Thus, the probability of detection must now be written as \( p(d,a) \). The probability of detection is still an increasing function of the...
state’s detection efforts $d$. But it is now also a decreasing function of violators’ detection avoidance spending, $a$.\(^{14}\)

An equivalent way of saying the same thing is as follows: positive detection avoidance activity increases the public detection cost of generating any given level of detection probability. Given that $p(d,a)$ increases in $d$, to say that positive detection avoidance activity $a$ reduces the probability of detection $p(d,a)$ for any level of public detection effort $d$, is also to say that, given greater detection avoidance activity $a$, any (still attainable) level of detection probability $p$ requires greater public detection effort $d$.

3.2.2.1.2 SECOND EFFECT Second, and counter to the first effect, detection avoidance activities are costly for those who engage in them, and such costs constitute, in effect, an additional component of the effective sanction for the underlying violation. Thus, although detection avoidance activities reduce the probability that violations will be detected, as just noted, the resulting reduction in deterrence is mitigated by the violator’s detection avoidance costs.

Formally, if a potential violator chooses to violate the law and, accordingly, engages in detection avoidance activity at the level of $a$, her “effective expected sanction” is $p(d,a)S + a$. This expression, or rather its minimized value with respect to the potential violator’s violation-contingent choice of $a$, is what the potential violator compares to the private benefit $b$ of an underlying violation in determining whether to commit the violation. Accordingly, the “level of deterrence” provided by the system is no longer the expected sanction $p(d)S$, but rather the effective expected sanction $p(d,a^*)S + a^*$, where $a^*$ minimizes $p(d,a)S + a$.

3.2.2.1.3 FIRST AND SECOND EFFECTS COMPARED It has been observed that the first and second effects just described are countervailing. It is thus worth asking: how do they compare in magnitude? This question must be answered separately with respect to (a) the total effect of detection avoidance activities on deterrence and (b) the marginal effect of such activities on deterrence.

The total effect of the violator’s optimally chosen detection avoidance activity on the level of deterrence can only be to lower her effective expected sanction for the underlying violation. That is, the dollar value of

---

\(^{14}\) In what follows I shall assume that $p$ is twice continuously differentiable, $p_a < 0$, $p_d > 0$, $p_{aa} > 0$, $\lim_{a \to 0} p_a = -\infty$, $\lim_{a \to \infty} p_a = 0$. These assumptions guarantee that $p(d,a)S + a$ is minimized at a strictly positive value of $a$, and that there is only one such minimizing value of $a$. 

---
the reduction in the probability of detection (the first effect identified in this subsection) must be at least as great as what the violator is spending on that reduction (the second effect identified in this subsection). Were this not true, the violator could, and would, improve her expected payoffs by engaging in no detection avoidance.

Formulaically, if $a^*$ is the violator’s optimal choice of detection avoidance activity, then $p(d,0)S - p(d,a^*)S$, the total reduction in the expected sanction due to privately optimal detection avoidance activity, must be at least as great as the private cost of detection avoidance $a^*$. This is the same as saying that $p(d,0)S$ must be at least as great as $p(d,a^*)S + a^*$. Were $p(d,0)S$ strictly less than $p(d,a^*)S + a^*$, the violator could do better by choosing $a = 0$.

By contrast, marginal changes from an optimally chosen amount of detection avoidance expenditure have no impact on the effective sanction for the underlying violation. This follows from the fact that an optimal amount of detection avoidance activity must satisfy the first-order condition for an interior minimum of $p(d,a^*)S + a^*$ with respect to $a$: $p_a(d,a^*)S + 1 = 0$. That is, the chosen level of detection avoidance cannot be optimal if it is possible to lower marginally the effective sanction $p(d,a)S + a$ by marginally increasing $a$ (as when $p_a(d,a^*)S + 1 > 0$) or by marginally decreasing $a$ (as when $p_a(d,a^*)S + 1 < 0$). Thus, on the margin, the amount by which additional detection avoidance reduces the expected sanction $pS$ is precisely equal to its cost to the violator, and the two countervailing effects on deterrence are precisely offsetting.

If marginals add to totals, how is a zero marginal reduction compatible with a positive total reduction? The claim is not that the marginal effect of detection avoidance on the effective expected sanction is zero everywhere, only that it is zero at the privately optimal level of detection avoidance. As detection avoidance is increased “step-by-step” from zero up to the optimal level, the marginal impact of each “step” will generally be to reduce positively the effective expected sanction. This can be shown formally using the fundamental lemma of calculus. Although $p_a(d,a)S + 1 = 0$ when $a = a^*$,

$$\int_{0}^{a^*} (p_a(d, \bar{a})S + 1)d\bar{a} = (p(d, a^*)S + a^*) - p(d, 0)S,$$

is negative.

3.2.2.2 POLICIES DIRECTED AT DETECTION AVOIDANCE ITSELF

3.2.2.2.1 THIRD EFFECT Third, adding detection avoidance to the story raises the possibility that detection avoidance itself can be sanctioned (a
Detection avoidance and enforcement theory

possibility, oddly, that is ignored by a fair portion of the small literature on detection avoidance.

Such “second-order sanctions” – on detection avoidance, rather than on the underlying violation – increase the level of deterrence for the underlying violation. As with the direct costs a of detection avoidance activities, whose effect was discussed in Section 3.2.2.1 above, this legally constructed cost counteracts the fact that detection avoidance reduces the detection probability for the underlying violation.

Thus, let $S_1$ be the sanction for detection avoidance. Let $p_1(a)$ be the probability that detection avoidance activity $a$ will be detected, an increasing function of $a$.\(^{15}\) (And, for the moment, assume that the detection of “first-order” detection avoidance cannot itself be avoided with “second-order” detection avoidance – an issue discussed in detail below in Section 6.) Then the effective sanction for underlying violations, and the level of deterrence for such underlying violations, is $p(d,a^*)S + a^* + p_1(a^*)S_1$, where $a^*$ is chosen to minimize this triple sum. For reasons described in the next paragraph, the derivative of this expression with respect to the sanction $S_1$ for detection avoidance activities is simply $p_1(a^*) > 0$. Therefore, increasing $S_1$ increases the level of deterrence for the underlying activity.

How can the derivative of $p(d,a^*)S + a^* + p_1(a^*)S_1$ with respect to $S_1$ be simply $p_1(a^*)$, when $S_1$ also affects $a^*$, and $a^*$ is a separate argument in $p(d,a^*)S + a^* + p_1(a^*)S_1$? Marginal changes in $S_1$ will indeed inspire marginal changes in detection avoidance $a^*$. But the effect of this responsive change in detection avoidance on the level of deterrence, $p(d,a^*)S + a^* + p_1(a^*)S_1$ will be nil. The reasoning is the same as in Section 3.2.2.1.3 above: if changes in the privately optimal level of detection avoidance did have a net marginal effect on the effective expected sanction $p(d,a^*)S + a^* + p_1(a^*)S_1$, the current level of detection avoidance would not be optimal for the violator. The proposition that one can ignore responsive behavioral changes when judging the impact on optimized payoffs of marginal parameter changes is referred to as the “envelope theorem.”

3.2.2.2 FOURTH EFFECT Fourth, as discussed in more detail in Section 7, accounting for detection avoidance turns government detection effort itself into a more complex policy variable. Accordingly, the range of alternative detection policies expands. With detection avoidance in the model, violators themselves can affect the probability of detection, at cost. Thus, given the government’s choice of detection activity $d$, the government

\(^{15}\) I shall assume that $p_1$ is twice continuously differentiable, $p_{1a}>0$, $p_{1aa}<0$, $\lim_{a^* \to 0} p_{1a} = \infty$, $\lim_{a^* \to \infty} p_{1a} = 0$. 

Gerrit De Geest - 10.4337/9781782540472.00012
Downloaded from Elgar Online at 11/26/2018 04:35:57PM via free access
effectively presents to the violator not a single probability of detection, \( p(d) \), but a “menu” of detection probabilities, \( p(d, a) \) one for each of her possible choices \( a \) of detection avoidance intensity. Section 7 describes how, by adjusting the shape of this menu, the state can affect both violators’ detection avoidance choices and their decision whether to violate the law in the first place.

### 3.2.3 The effect of detection avoidance on social cost accounting

Adding detection avoidance to the model alters not only the deterrence mechanics of the conventional enforcement model, but also its social cost-benefit accounting. Recall from Section 3.1.2 that the conventional approach to enforcement, which ignores detection avoidance, focuses on the publicly incurred cost of “detecting” underlying violations, as well as the direct social cost of sanctions like imprisonment. To these two costs, three new costs must be added. These new costs fall naturally into two groups.

#### 3.2.3.1 The private cost of detection avoidance

The first necessary addition to social cost-benefit accounting is the private cost of detection avoidance: expenses incurred by private parties in hampering investigation and fighting prosecution. In the model laid out above (in which, implicitly, violators differ only in their private benefits from underlying violations), every violator engages in the same level of detection avoidance: that which minimizes \( p(d,a^*)S + a^* + p_1(a^*)S_1 \). The number of violators is thus \( N(1 - F(p(d,a^*)S + a^* + p_1(a^*)S_1)) \). Detection avoidance costs are thus \( N(1 - F(p(d,a^*)S + a^* + p_1(a^*)S_1))a^* \).

#### 3.2.3.2 The costs of sanctioning detection avoidance

The other two costs that must be added to the analysis arise to the extent that detection avoidance is itself subject to sanction. They are merely the counterparts to the two costs – of detection and of sanctioning – that are emphasized in the conventional paradigm.

First, there is the public cost of detecting detection avoidance. Perjury, for example, must also be investigated and prosecuted. Thus, the probability that detection avoidance is detected is not actually \( p_1(a^*) \), as above, but rather \( p_1(d_1,a^*) \), where \( d_1 \) is the public cost of detecting detection avoidance.

Second, there is \( c(S_1) \), the direct cost of imposing the sanction \( S_1 \) on detection avoidance, such as the costs of imprisonment.

---

16 See notes 6 and 13: in this model, non-violators do not engage in detection avoidance because they are never wrongfully sanctioned.
4 The Effect on Detection Avoidance Activity of Sanctioning the Underlying Activity

It was noted above that increasing the sanction on detection avoidance increases the level of deterrence for the underlying violation and thus reduces commission of the underlying violation. An important segment of the literature on detection avoidance focuses on the reverse direction of causation: the effect of sanctioning the underlying activity on the incidence of detection avoidance.

Malik (1990), which is perhaps the first systematic account of detection avoidance in the context of the conventional economic enforcement model, recognizes that raising the sanction on the underlying activity inspires the expenditure of additional private resources on detection avoidance. (See also Snyder 1990.) When the government raises a fine, for instance, it increases not only the pain of detection for the violator, but also her relief from avoiding detection. Were the fine $100,000, a detection avoidance activity that reduced the chance of detection by one percentage point would be worth $1000 to the violator. Doubling the fine to $200,000 doubles the value of that activity to $2000.

That increasing the sanction on the underlying activity increases detection avoidance can be seen in terms of the model introduced in Section 3.2. Assuming that detection avoidance is not itself sanctioned (as does Malik 1990), the violator’s chosen level of detection avoidance satisfies the first-order condition $p_a(d_a*)S + 1 = 0$, where, the reader will recall, $p_a < 0$. Applying the implicit function theorem to this first-order condition, we see that the marginal impact on detection avoidance of raising the sanction on the underlying violation is $\frac{da*}{dS} = -\frac{p_a}{Sp_{aav}}$. This marginal impact is positive because detection avoidance reduces the probability of detection.

---

17 On close inspection, it appears that Tabbach and Nussim (2008) are making essentially the same point when they show how an increase in “ex post punishment of detection avoidance” may increase avoidance activities. In their model, an increase in the level of “ex post punishment for detection avoidance” also increases the sanction for the underlying activity.

18 The (single-dimensional) implicit function theorem – which is applied repeatedly to first-order conditions throughout this chapter – says, roughly, that, if $F(x, y) = 0$, $(dx/dy)_{F=0} = -F_y/F_x$. The derivative $(dx/dy)_{F=0}$ is the change in $x$ that, per change in $y$, keeps $F$ constant at 0. Intuitively, if $x$ increases $F$ at a rate of 2 to 1 ($F_x = 2$), and $y$ increases $F$ at a rate of 1 to 1 ($F_y = 1$), then decreasing $x$ at half the rate at which $y$ is increased keeps $F$ constant. Among other things, the theorem assumes the condition $F_x \neq 0$, a condition that is usually satisfied in the context of first-order conditions by assuming the strict concavity (convexity) of the function to be maximized (minimized).
Procedural law and economics

\( p_a < 0 \) at a decreasing rate \( p_{aa} > 0 \).\(^{19}\) Therefore, raising the sanction for the underlying activity raises the violator’s chosen level of detection avoidance.\(^{20}\)

Let us now consider the impact of this dynamic – first on deterrence mechanics, and then on social cost-benefit accounting.

4.1 The Effect of Responsive Increases in Detection Avoidance on Deterrence Mechanics

Given that raising the sanction for the underlying violation induces more detection avoidance, does this imply that raising the underlying sanction can actually induce more underlying violations? On an intuitive level, it might seem that the responsive increase in detection avoidance could so lower the probability of detection \( p(d,a) \) as to outweigh the impact of raising the ex post sanction \( S \) – thus causing the expected sanction to fall and inducing more underlying violations.

This intuition is problematic, however. It fails to account for the fact that increasing detection avoidance is costly for the violator. It specifically ignores the fact that, as described above in Section 3.2.2, marginal changes in detection avoidance away from the privately optimal level thereof have no marginal impact on the effective expected sanction because the reduction in the probability of detection is precisely off set by the increase in private detection avoidance costs.

Formally, the effective expected sanction for the underlying activity is \( p(d,a^*)S + a^* \) (given that detection avoidance itself is not sanctioned). By the envelope theorem,\(^{21}\) the derivative \( \frac{da^*}{dS} \) of this expression with respect to \( S \) is simply \( p(d,a^*) \), which is positive. Thus, increasing \( S \) does indeed increase the effective expected sanction for the underlying violation, thus lowering its incidence – all this despite the responsive increase in detection avoidance.

The problematic intuition described above focuses on the change in \( p(d,a^*)S \) due to changes in \( a^* \) that are inspired by changes in \( S \). But the effective expected sanction is \( p(d,a^*)S + a^* \) not \( p(d,a^*)S \). And the effect on the effective expected sanction of the responsive change in detection avoidance is \( p_a(d,a^*)S + 1 \), which, the reader will recall, must be zero at optimally chosen \( a^* \).

---

\(^{19}\) See note 14. Note that a necessary condition for a minimum is \( p_{aa} \geq 0 \). Therefore, \( da^*/dS \) can never be finite negative or negative infinity, even if the assumption \( p_{aa} > 0 \) is abandoned.

\(^{20}\) See note 14.

\(^{21}\) See Section 3.2.2.2.1 above.
4.2 Extension: The Effect of Responsive Increases in Detection Avoidance on Deterrence Mechanics when Violation is a Continuous Variable

The same principle applies if the underlying violation is not a binary variable, but rather a continuous variable, so that there are varying degrees of violation. If the violator optimizes detection avoidance based in part on her chosen level of underlying violation, then responsive adjustments in detection avoidance due to changes in the sanction for that level of underlying violation have no impact on the effective expected sanction for that level of underlying violation.

The model can be (temporarily) extended to make this point formally. Let detection effort \( d \) be fixed and implicit. Let \( v \) be the level of underlying violation. Let \( S(v,B) \) be a violation-contingent sanction, where \( B \) is a parameter that we shall use to adjust the sanction as a function of \( v \). Let \( a^*(v,B) \) be optimal violation-contingent detection avoidance for the violator. Then \( p(v,a^*(v,B))S(v,B) + a^*(v,B) \) is the effective expected sanction for the underlying violation level \( v \).

For any given level of underlying violation, \( v \), the marginal change in the effective expected sanction with respect to sanctioning parameter \( B \) is, by the envelope theorem, \( p(v,a^*(v,B))S_B(v,B) \). Just as in the binary violation model discussed in the preceding section, responsive changes in detection avoidance, \( a^*(v,B) \), do not appear directly in this expression. Responsive changes in detection avoidance do not directly affect the marginal impact of \( B \) on the effective expected sanction for underlying violation level \( v \). 22

This basic point must be kept in mind when interpreting Nussim and Tabbach (2009), a recent contribution on this issue. Nussim and Tabbach’s main assertion is that, in a model with detection avoidance, increasing the underlying sanction can actually induce an increase in the level of underlying violations due to responsive increases in detection avoidance. In a representative passage, they say:

[H]arsher punishments may . . . result in more crime. . . . [C]onsider how a small additive, constant increase in punishment across all levels of criminal activities will affect the behavior of offenders . . . Such an increase will have no direct effect on the choice of crime, since it does not alter marginal expected punishment. Yet, since total punishment is now higher, the marginal benefit from avoidance is greater, thus inducing offenders to invest more in avoidance. But once offenders invest more in avoidance, the probability of punishment falls, and

22 Again, this is a result of the fact that the marginal impact of already optimized detection avoidance on the expected sanction \( p_d(v,a^*(v,B))S(v,B) \) must balance with its marginal cost, which is 1. That is, the fact that \( p_d(v,a^*(v,B))S(v,B) + 1 = 0 \) implies that the derivative of \( p(v,a^*(v,B))S(v,B) + a^*(v,B) \) in \( B \) via \( B \)'s inclusion in \( a^*(v,B) \) (i.e., holding \( S(v,B) \) constant) is zero.
marginal expected sanctions [i.e., the derivative of expected sanction with respect to the level of the underlying violation] fall as well. This complementary effect increases crime levels.23

It appears that Nussim and Tabbach may have misconstrued their model. Using the notation introduced above, the effective expected sanction for violation level \( v \) in Nussim and Tabbach’s hypothetical is

\[
p(v,a^*(v,B))(S(v) + B) + a^*(v,B).
\]

Notice that in this expression \( B \) is indeed a “constant increase in punishment across levels” of the underlying violation. Notice also that \( a^*_{g}(v,B) \) is indeed positive – that is, it is true that “since total punishment is . . . higher [when \( B \) is increased], the marginal benefit from avoidance is greater, thus inducing offenders to invest more in avoidance.” And because \( p_a < 0 \), it is indeed true that “once offenders invest more in avoidance, the probability of punishment falls.” However, it would be a non sequitur to conclude from this, as Nussim and Tabbach seem to, that “marginal expected sanctions fall as well.” By the envelope theorem-based reasoning described above, the responsive change in detection avoidance inspired by the increase in \( B \) has no direct impact on the effective expected sanction at each level of \( v \), \( p(v,a^*(v,B))(S(v) + B) + a^*(v,B) \). Therefore, the response change in detection avoidance also has no direct impact on “marginal expected sanctions:” that is, it also has no impact on the derivative of \( p(v,a^*(v,B))(S(v) + B) + a^*(v,B) \) in \( v \).

To more clearly identify the logical problem, consider the following special case. Suppose that sanctions are initially uniform, \( S(v) = S \). Assume, also (like Nussim and Tabbach) that the probability of detection \( p \) is not a function of the level of the violation \( v \), so that \( p(v,a^*(v,B)) = p(a^*(v,B)) \). Given these two assumptions, contingently optimal detection avoidance, and so the probability of detection, are the same across all possible choices for \( v \). That is, for all \( v \), \( a^*(v,B) = a^*(B) \) and \( p(a^*(v,B)) = p(a^*(B)) \). Increasing \( B \) still increases detection avoidance – \( a^*_{g}(B) \) is still positive. Thus Nussim and Tabbach’s premise holds. However, Nussim and Tabbach’s conclusion does not; increasing \( B \) provides no additional inducement to increase the level of the violation. The effective expected sanction was formerly the same at every \( v \), namely \( p(a^*(B = 0))S + a^*(B = 0) \). After increasing \( B \), it is still the same at every \( v \), namely \( p(a^*(B))S + a^*(B) \). Thus the marginal effective expected sanction was and is zero.

This is not to deny that increasing the underlying sanction may inspire a greater level of the underlying violation. But this is for different reasons than Nussim and Tabbach provide. Returning to the setting in the

paragraph before last, the change in the effective sanction at each level of $v$ is, by the envelope theorem, simply $p(v, a^*(v, B))$. Thus, when sanctions are increased by the same amount across the board, the effective expected sanction at any given $v$ increases according to the initial probability of detection at $v$. As a result, the increase in the expected effective sanction will be greater at levels of the violation that are initially more likely to be detected. If, in particular, greater levels of the underlying violation are more likely to be detected, then the uniform increase in the sanction will increase the sanction by a greater amount at greater levels of the underlying violation and so induce a decrease in the level of the underlying violation. If, on the other hand, lower levels of the underlying violation are more likely to be detected, then the uniform increase in the sanction will induce an increase in the level of the underlying violation – thus generating the perverse result that Nussim and Tabbach highlight.

What does this have to do with detection avoidance? In one sense, it has nothing to do with detection avoidance. More precisely, the perverse result that Nussim and Tabbach represent to be a consequence of detection avoidance may arise in a model without detection avoidance. In such a model, the increase in the expected effective sanction at $v$ due to increasing $B$ would again be the probability of detection at $v$, now simply $p(v)$. Thus the increase in the expected sanction would be greater for lower levels of $v$, if lower levels of $v$ were more likely to be detected.

Yet the question then arises: why would the probability of detection ever be greater for lower violations than for greater violations? And here detection avoidance activity may have a role to play. Suppose (as do Nussim and Tabbach) that the probability of detection is solely a function of detection avoidance. Assume, not unreasonably, that initial sanctions are increasing in the level of the violation. Then contingently optimal detection avoidance is greater for higher levels of the underlying violation than for lower. Accordingly, the probability of detection – assumed to be a function solely of detection avoidance – is greater for lower levels of the underlying violation than for higher. Therefore, a uniform increase in the sanction induces an increase in the underlying violation.

There are, however, two caveats, both having to do with the fact that the probability of detection is not in fact solely a function of detection avoidance. First, the level $v$ of the underlying violation probability has a direct impact on the probability of detection. It is not implausible to suppose that greater levels of the underlying violation emit greater amounts of evidence, leading, all else the same, to a $v$-increasing detection probability. Second, the probability of detection is also a function of the government’s detection effort: detection is not just a flight, it is also a pursuit. It is not implausible that authorities will investigate and prosecute
larger violations with greater vigor, again leading, all else the same, to a $v$-increasing detection probability. In the end, therefore, the question is how the combination of these three factors – variation in detection effort, variation in detection avoidance effort, and variation in the nature of the violation – determines the variation in detection probabilities across levels of underlying violation.24

A second point about Nussim and Tabbach’s (2009) analysis concerns the structure of the ex post sanction. The last several paragraphs have considered the case in which the increase in the sanction is the same at every level of $v$. This was solely for analytical convenience. The authorities might instead increase the ex post sanction to a greater extent for larger violations. Doing so could eliminate the perverse effect that Nussim and Tabbach highlight. Returning to the more general analysis at the beginning of this section, the change in the effective expected sanction at any given $v$ due to any increase in $B$ is $p(v,a^{*}(v,B))S_{p}(v,B)$. Because $S_{p}(v,B)$ is generally a function of $v$, the relative impact on the effective expected sanction at high and low underlying violations is under the control of the authority setting the sanctioning structure. Even if $p(v,a^{*}(v,B))$ were decreasing in $v$, the authority could structure the sanction increase so that $S_{p}(v,B)$ increased in $v$ at a rate that exceeded the rate at which $p(v,a^{*}(v,B))$ decreased in $v$. This would be sufficient to guarantee that the product $p(v,a^{*}(v,B))S_{p}(v,B)$ increased in $v$. And this would mean that the $v$-increasing increase in the sanction induced a reduction in the level of the underlying violation.

What then is the takeaway point from Nussim and Tabbach (2009)? Perhaps it is this: the existence of detection avoidance is most likely to make $p(v,a^{*}(v,B))$ less $v$-increasing than it would otherwise be, and possibly even $v$-decreasing. Consequently, an increase in the ex post sanction that would otherwise be sufficiently $v$-increasing to induce a reduction in the level of the underlying violation may not be so by virtue of detection avoidance activity. Nussim and Tabbach (2009) provide an example in which the ex post sanction is $S(v,B) = Bv$, as if $B$ were the price of crime. In this case, the marginal increase in the sanction at $v$ due to an increase in $B$ is $v$ itself, and so is increasing in $v$. Nussim and Tabbach then provide a condition on the probability of detection function $p(a)$ under which this manner of increasing the sanction would still induce an increase in the violation.

---

24 Both of the effects identified in this paragraph also call into question the joint assumption in the previous paragraph that detection avoidance activity, all else the same, increases with the level of the underlying violation and acts to make the probability of detection greater for lower levels of underlying violation.
underlying violation. They also identify a specific probability function, based on the logarithm, which would fulfill their condition.25

4.3 The Effect of Responsive Increases in Detection Avoidance on Social Cost-Benefit Accounting

The responsive change in detection avoidance has important implications for social cost-benefit accounting. The main point of Malik (1990) itself is that this dynamic effectively adds an additional component to the cost of underlying sanctions, and thus generates (yet another) qualification to Becker’s (1968) famous prescription – crafted in a model without detection avoidance – that it is generally best to generate any desired level of deterrence with large monetary fines and small detection probabilities. Several subsequent contributions, discussed below, have qualified Malik’s qualification. As elsewhere in enforcement theory,26 Becker’s prescription has been the chief concern of most research on detection avoidance.

4.3.1 Malik’s qualification to Becker

Under Becker’s analysis (much simplified), a monetary fine is merely a transfer of resources from the offender to the government and a virtually costless means of generating additional deterrence. That is, if \( S \) is a fine, \( c(S) = 0 \) for all \( S \). By contrast, increasing the chance that violations are detected diverts labor and capital away from other productive activities. Best then, concludes Becker, to lower detection effort, only rarely catching offenders, and to compensate by imposing large fines upon those few who are caught. That is, any level of deterrence \( p(d)S \) may be generated with lower public costs by lowering...

---

25 Nussim and Tabbach (2009: 317). Nussim and Tabbach draw this example and its analysis from a referee report received on an earlier draft.

26 The literature qualifying Becker’s prescription is extensive and begins shortly after Becker’s article was published in 1968. For example, Stigler (1970) focuses on cross-offense incentives and argues that if all fines are set to the same maximum level, offenders will choose serious rather than minor offenses, at least if detection probabilities cannot be appropriately adjusted across offenses. Polinsky and Shavell (1979) incorporate additional risk-bearing costs borne by risk-averse offenders when sanctions are increased. Bebchuk and Kaplow (1992) find that increasing sanction multiplies the effect of individuals’ errors in judging probability of detection and thus exacerbates over- and underdeterrence. Bebchuk and Kaplow (1993) find that lowering the fine and perhaps raising the detection probability facilitates imposition of effectively separate expected sanctions according to individuals’ heterogeneous ability to avoid detection, thus preventing over- or underdeterrence. Kahan (1997) notes that “if individuals infer widespread criminality from a low probability of apprehension, the power of social influence could more than offset any efficiency gains from this tradeoff.” See also the list of qualifications described in the surveys listed in note 4.
d (thus reducing public detection costs) and raising the level of the fine S (which has no impact on public sanctioning costs) by enough to make up for the reduction in p(d).27

Yet, as Malik (1990) points out, although raising the fine may not incur the expenditure of additional public resources on detection, it does inspire the expenditure of additional private resources on detection avoidance. Consequently, monetary fines and detection probabilities are placed on more equal footing with respect to how efficiently they generate deterrence, and one can no longer claim that the monetary fines should carry all (or even more) of the weight.

Formally, the level of deterrence, which is the level of the effective expected sanction, is \( p(d,a^*(S))S + a^*(S) \), where it is implicitly assumed that \( p_{da} = 0 \), so that the violator’s level of detection avoidance is a function solely of the ex post sanction. If S is a fine, so that \( c(S) = 0 \), then the expected social cost of generating this effective expected sanction is \( a^*(S) + d \). Now repeat Becker’s maneuver: raise S and lower d so as to keep the level of deterrence, \( p(d,a^*(S))S + a^*(S) \) the same (which now requires accounting for the sanctioning role of detection avoidance). In Becker’s model, in which there is no detection avoidance, social costs are simply d, and the maneuver lowers such costs by decreasing d. In Malik’s model, the increase in S causes an increase in \( a^*(S) \), and it is not generally possible to say whether the sum \( a^*(S) + d \) increases or decreases.

4.3.2 Innes’ qualification to Malik Innes (2001) shows that Becker’s prescription is restored – even in the presence of costly detection avoidance – if the government invites violators to self-report and charges them a self-reporting sanction equal to the effective expected sanction for non-self-reporting. That is (assuming again that detection avoidance is not sanctioned), the government sets the sanction for self-reported violations equal to (a small amount less than) \( p(d,a^*)S + a^* \), and this induces all violators to self-report. The result is that deterrence is maintained and no detection avoidance costs are actually incurred. Therefore, actually incurred detection avoidance costs (which are zero) are not sensitive to the combination of d and S which is used to create the (never carried out) threat of \( p(d,a^*)S + a^* \). (In Innes’ model, even with self-reporting, the

27 Becker also finds that even if the sanction is socially costly to impose (contrary to the monetary fines considered in this paragraph), it will still be efficient to increase the sanction and lower the detection probability if the social cost elasticity of sanctions is no greater than one. A similar issue arises in Tabbach (2010) discussed below in Section 4.3.5. For more on the conventional model of enforcement, see Garoupa (1997), Polinsky and Shavell (2007), and Franzoni (2000).
government must still actually incur public detection costs \( d \). It follows that it is optimal to create such threat in a manner that minimizes the sum of \( d \) and \( c(S) \), just as in Becker (1968). Thus, Becker’s prescription is restored.

As Innes notes, this stark result depends, *inter alia*, on an assumed uniformity across violators of \( p(d,a^*)S + a^* \), the effective expected sanction for non-reporting. In his model, as in the model that is being carried along in this entry, violators differ only in their private benefits \( b \) for violating the law, and not in any other way. In particular, violators do not differ in their detection avoidance opportunities or efficiencies. If violators did so differ, and such differences were not fully observable by the government, any self-reporting sanction structure would generate infra-marginal detection avoiders – violators who choose not to self-report and who, accordingly, actually incur detection avoidance costs. In this case, detection avoidance costs would again be a factor in determining the optimal enforcement structure.

Arguably, Innes’ assumption that violators are homogeneous in detection avoidance creates a problematic asymmetry in his analysis. On the one hand, the detection avoidance problem itself is in large measure created by the fact that, on the level of the underlying violation, it is assumed that there is sufficient heterogeneity to ensure the existence of infra-marginal violators. These infra-marginal violators are the ones whom we would expect to incur the greatest detection avoidance costs. On the other hand, the solution to the detection avoidance problem that Innes studies is in large measure dependent on the fact that, on the level of detection avoidance, there is sufficient homogeneity to ensure the non-existence of infra-marginal detection avoiders. Presumably, the general principle that justifies heterogeneity on the level of the underlying violation also justifies heterogeneity on the level of detection avoidance.

Nonetheless, we may take from Innes (2001) the following general lesson: self-reporting can economize not only detection costs (as was already known), but also, and for similar reasons, detection avoidance costs.

### 4.3.3 Langlais’ first qualification to Malik

Langlais (2008) considers the case in which public detection effort and private detection avoidance effort are “strategic complements:” when the authorities increase the level \( d \) of detection, violators are inspired to increase the level of their detection avoidance. This phenomenon arises when \( p_{da} < 0 \), so that the change in the probability of detection brought about by an increase in detection avoidance, \( p_a < 0 \), is of greater (negative) magnitude the greater is public detection effort. The situation is not implausible, at least over some ranges.
Consider that if there is no detection effort at all, then detection avoidance is irrelevant: the probability of detection is zero in any event. Likewise, if the authorities have devoted sufficient resources to check every file, then there is a greater expected benefit to the violator herself from going through each and every file, destroying that which is incriminating.

Now consider the Beckerian maneuver of raising the fine and lowering the detection probability. Raising the fine inspires an increase in costly detection avoidance, as Malik (1990) emphasizes in arguing such maneuver is not always beneficial. However, when detection and detection avoidance are strategic complements, the simultaneous reduction in detection effort inspires violators to reduce detection avoidance. If this reduction in detection avoidance caused by the reduction in detection exceeds the increase in detection avoidance caused by the increase in the fine, then the beneficial character of Becker's maneuver is restored.\textsuperscript{28}

### 4.3.4 Langlais' second qualification to Malik

Langlais (2009) considers a different kind of violator activity, which he calls “dissembling.” Dissembling refers to efforts by violators to prevent the disgorgement of illegal benefits. In Langlais’ model, violators who are caught must disgorge their apparent illicit benefits and pay an additional “punitive” fine. Violators may choose to expend effort dissembling in order to hide their benefits, and these efforts are socially costly. Increasing the punitive portion of the fine in this model does not induce additional dissembling effort, because the benefits of dissembling do not depend on the punitive part of the sanction. Accordingly, Langlais finds, in his central case, that the punitive fine should be maximal, as in Becker. The upshot is that not all types of avoidance activities are the same, and some do not lead to the effects highlighted by Malik (1990).

### 4.3.5 Tabbach’s qualification to Malik

Malik (1990) studies the case of fines, wherein the social cost of imposing ex post punishment $S$ is zero: $c(S) = 0$. In contrast, Tabbach (2010) studies the case in which the social cost of imposing ex post punishment $S$ exceeds the punishment itself: $c(S) > S$.\textsuperscript{29} Tabbach’s case might arise where the ex post punishment is generated solely by imprisonment and imprisonment incurs social costs over and above those borne by the imprisoned violator. Tabbach specifically considers the subcase in which $c(S)$ is linear, taking the form $cS$ so that $c$ is

\textsuperscript{28} Langlais (2008) also presents several results concerning changes in the maximal fine in the case that the maximal fine is optimal (pp. 377–8).

\textsuperscript{29} In Tabbach’s notation, $S$ is “$\alpha S$” and $c(S)$ is “$(\alpha + \gamma)S$.”
the social “price” of imposing a dollar’s worth of ex post sanction on the violator. Tabbach shows (in effect) that when $c > 1$, Becker’s prescription to maximize $S$ is restored despite detection avoidance costs.

One way to present the argument is this. The effective expected sanction is, as in Malik (1990), $p(d, a^*(S))S + a^*(S)$. But expected social costs are now $p(d, a^*)cS + a^* + d = cp(d, a^*)S + a^* + d$. Consider the following modified Becker maneuver: raise $S$ and lower $d$ in such manner that the effective expected sanction $p(d, a^*)S + a^*$ remains constant. The maneuver causes $a^*$ to increase (by raising $S$). Thus, the maneuver must cause $p(d, a^*)S$ to decrease by an equal amount (by lowering $d$ sufficiently). Therefore – and this is the key step – given that $c > 1$, the maneuver causes $cp(d, a^*)S$ to decrease by more than $a^*(S)$ increases. Therefore, $cp(d, a^*)S + a^*$ decreases. Thus, expected social costs, $cp(d, a^*)S + a^* + d$, decrease: the maneuver decreases not only detection costs $d$, but also “non-detection expected social costs” $a^* + cp(d, a^*)S$.

If, on the other hand, the social price of imposing a dollar of ex post sanction on the violator was any amount less than $\$1$, then, in the key step, $cp(d, a^*)S$ would decrease by less than $a^*(S)$ increases. Consequently, although Becker’s maneuver would still decrease $d$, it would increase non-detection expected social costs, $cp(d, a^*)S + a^*$.

The intuition for these results may be stated as follows:

1. Ignore detection costs $d$, and focus on non-detection expected social costs, $a^* + cp(d, a^*)S$. In performing Becker’s maneuver – which effectively raises $a^*$ while keeping $p(d, a^*)S + a^*$ constant – society is effectively replacing dollars of effective expected sanction in the form of $p(d, a^*)S$ with dollars of effective expected sanction in the form of detection avoidance costs $a^*$. The social cost of creating effective expected sanctions with detection avoidance costs is one-to-one. If the social cost of creating effective expected sanctions through $p(d, a^*)S$ is anything greater than one-to-one – if $c > 1$ – then the substitution effected by the Becker maneuver reduces non-detection expected social costs. If, on the other hand, the social cost

---

30 As in Malik (1990) and Tabbach (2010), the assumption is that $p_{ds} = 0$, so that $a^*$ is a function solely of $S$.

31 This case is largely implicit in Tabbach (2010). See, however, the last paragraph of the article’s text (p. 285).

32 The following intuition is provided in Tabbach (2010: 268–9): “The reason is simple. Without avoidance, increasing imprisonment and reducing its probability so as to leave the expected sanction unaltered saves on enforcement costs without affecting the costs of punishment. With avoidance, increasing imprisonment provides another type of benefit because it induces more avoidance, which substitutes for costlier punishment, thereby saving not only enforcement but also punishment costs.”
of creating effective sanctions through \( p(d,a^*)S \) is anything less than one-to-one – if \( c < 1 \) – then the substitution effected by Becker’s maneuver increases non-detection expected social costs.

Some remarks on the condition \( c > 1 \) are in order.

First, the condition is sufficient for restoration of Becker’s maximum sanction result, but not necessary, since detection costs \( d \) also decrease.

Second, despite some indications to the contrary in the introduction to Tabbach (2010), \( c > 1 \) does not automatically hold whenever sanctions fail to be purely monetary. Indeed, mixing in a small fine – small relative to the size of the additional social cost of non-monetary fines – can cause the condition to fail. This is because fining delivers dollars of effective expected sanction at a cost of zero-to-one, and so mixing fines into \( S \) quickly brings down the average cost of creating effective expected sanctions through \( S \).

Suppose, for example, that 40% of the dollar value to the violators of ex post sanctions \( S \) came in the form of monetary transfers to the government (or to plaintiffs) and 60% came in the form of imprisonment. Suppose further that each dollar of the portion of the ex post sanction that is generated by imprisonment (such portion being 0.6\( S \)) costs society $1.50 in addition to the cost, $1, borne by the violator. Then the cost \( c(S) \) of the ex post sanction \( S \) would be the violator’s cost of imprisonment 0.6\( S \) plus the society’s additional cost of imprisonment 1.5(0.6\( S \)). This totals 0.9\( S \). Thus, \( c \) would equal 0.9, and it would not be possible to conclude that Becker’s maneuver lowered non-detection expected social costs.

More generally, if \( f \) is the fraction of \( S \) generated by monetary sanctions, and \( m \) is the additional social cost per dollar value of imprisonment cost to the violator, then \( c = f0 + (1-f)(1+m) \). Thus, \( c \leq 1 \) – and the condition fails – so long as the proportion of the total ex post sanction that is monetary, \( f \), is greater than or equal to the proportion of the total social cost of imprisonment that is not borne by the violator, \( ml/(1 + m) \). Notice that, as the latter proportion, \( m \), goes to zero, so does the threshold fraction of \( f \) for the condition \( c \leq 1 \).

33 According to the introduction of Tabbach (2010), “Th[e] article shows that Malik’s (1990) argument against maximal sanctions applies if and only if sanctions are monetary or more generally have no social costs above and beyond the costs incurred by offenders” (p. 268). Further, we read that “if optimality requires utilizing fines and imprisonment, Malik’s (1990) argument again does not apply, whereas Becker’s (1968) result holds” (p. 269).

Compare these introductory statements to the last section, and in particular, the last text paragraph of the article (p. 285), which is consistent with the analysis in this entry.
Third, as Tabbach (2010) makes clear, the restoration of Becker’s result presumes that the government is constrained to use sanctions for which $c > 1$. If sanctions with $c < 1$ are available, optimization over sanctioning form will require using these instead, and Malik’s qualification will again apply.

Fourth, the relationship between imprisonment and the case that $c > 1$ may be less clear than it first appears. It certainly seems plausible that imprisonment requires government outlays over and above the dollar sanction borne by the violator. But, on the other hand, imprisonment may in some cases also generate social benefits in the form of incapacitation. These social benefits are not enjoyed by the violator, and so they act to reduce $m$ (in the notation above) and so $c$. Indeed, there might be cases in which $m < 0$ – that is, in which the social benefit of incapacitation exceeds the non-privately borne costs of imprisonment. In this case, $c < 1$ even if the fraction of monetary sanctions, $f$, is zero.

Fifth, it should be noted that despite the substantive law focus of the root article in this area, Becker (1968), the detection avoidance dynamic identified in Malik (1990) applies in the regulatory context and in civil litigation, where monetary fines are dominant.

Sixth, if the social cost of the ex post sanction $c(S)$ is nonlinear, the relevant condition is more complicated than the mere analogue to $c > 1$, which would be $c(S) > S$. The curvature of $c(S)$ also becomes important. Referencing the key point identified above, the fact that the maneuver causes $p(d,a^*)S$ to decrease by the same amount that $a^*$ increases, does not imply that $p(d,a^*)c(S)$ decreases by more than $a^*$ increases, even if $c(S) > S$. If $c(S)$ increases at an increasing rate, then a change that raised $S$ but left $p(d,a^*)S$ constant – which constitutes a mean-preserving spread of the distribution of the ex post sanction (viewed as a random variable) – would increase convex $c(S)$ (by Jensen’s inequality). This effect would work against the fact that, in Becker’s maneuver, $p(d,a^*)S$ is not held constant, but decreases.

5 Sanctioning Detection Avoidance – Assuming No Recursivity

5.1 The Importance of Considering Sanctions for Detection Avoidance

Several of the papers discussed in Section 4 immediately above – Malik (1990), Innes (2001), Nussim and Tabbach (2009), for example – assume that it is not possible to sanction detection avoidance itself. The

34 See, for example, the analysis of decoupling in Choi and Sanchirico (2004).
35 Indeed, Becker (1968) itself analyzes this issue.
36 Tabbach (2010) discusses sanctioning detection avoidance when $c > 1$ (see
assumption is occasionally supported by arguing that sanctioning detection avoidance is infeasible because, for example, such activities are unobservable by the enforcement authority. This defense is problematic both conceptually and empirically.

On a purely conceptual level, it would be difficult to argue successfully that detection avoidance activities are categorically more difficult for the government to detect and sanction than underlying violations. Each category houses a diverse variety of activities. Moreover, certain underlying violations (such as fraud, insider trading, slander, etc.) resemble detection avoidance, while certain detection avoidance activities (e.g., witness intimidation) resemble underlying violations.

On an empirical level, detection avoidance – in the form of lying to investigators, obstruction of justice, perjury, discovery misconduct, etc. – is in fact detected and sanctioned. Indeed, some commentators decry the fact that prosecutors will too often abandon prosecution of the underlying crime in favor of pursuing sanctions for various forms of derivative detection avoidance. It is said that prosecutors often do this because detection avoidance crimes are easier, not harder, to prosecute.

5.2 The Potentially Decisive Impact of Adding Sanctions for Detection Avoidance

As discussed in Section 4.3, Malik (1990) shows that accounting for detection avoidance can overturn Becker’s conventional prescription to generate deterrence via large fines and low detection probabilities. Several authors have since provided the literature with a list of conditions – concerning self-reporting, strategic complementarities, and sanctioning costs – that restore Becker’s prescription despite the dynamics of detection avoidance. However, the very possibility of sanctioning detection avoidance itself restores Becker’s prescription – at least if it is assumed, as it shall be in this section (but not in the next, Section 7), that the detection of detection avoidance activity cannot itself be avoided.
Recall that Becker’s prescription is to lower public detection costs \( d \) and raise the fine \( S \) in such manner as to maintain deterrence, the sole result being a reduction in social costs. Recall also Malik warning that raising the fine increases the private cost of detection avoidance. But the private cost of detection avoidance may be kept in check by simultaneously, and costlessly, raising the fine \( S_1 \) on detection avoidance.

Thus, after lowering public detection costs \( d \) à la Becker, instead of raising only the fine on the underlying violation, \( S \), let the government raise both \( S \) and \( S_1 \) together. Let it do so at relative rates that guarantee that the positive impact of raising \( S \) on detection avoidance is precisely cancelled by the negative impact on detection avoidance of raising \( S_1 \). (The next paragraph describes how such relative rates may be determined.) Let this relatively calibrated joint increase in \( S \) and \( S_1 \), which induces no change in detection avoidance activity, be itself calibrated to offset the reduction in deterrence of the underlying violation due to the reduction in \( d \). The end result is a reduction in public detection costs, no change in the level of deterrence for the underlying violation, and no increase in detection avoidance.

Formulaically, imagine that violators choose detection avoidance \( a \) to minimize an effective expected sanction that includes a sanction for detection avoidance: 

\[
p(d,a)S + a + p_1(a)S_1, \text{ as discussed in Section 3.2.3.1.}
\]

Then, per Section 4, the implicit function theorem may be applied to the first-order condition for the effective expected sanction-minimizing level of \( a^* \) in order to establish that the change in detection avoidance due to the increase in the fine on the underlying violation is 

\[
dS_1 = \frac{\partial p(d,a)}{\partial d} \frac{dS_1}{dS} > 0.
\]

Also by the implicit function theorem, the change in detection avoidance due to an increase in the fine on detection avoidance, is 

\[
dS_1 = \frac{\partial p_1(a)}{\partial S_1} \frac{dS_1}{dS} < 0.
\]

Thus, increasing the fine on the underlying violation increases detection avoidance, while increasing the fine on detection avoidance decreases it. Specifically, marginally increasing \( S \), while simultaneously marginally increasing \( S_1 \) at the relative rate of 

\[
\frac{dS_1}{dS} = -\frac{p_d}{p_{1a}} > 0,
\]

has, in total, no marginal impact on \( a^* \). Nevertheless, this coordinated increase in both sanctions causes the effective expected sanction for the underlying violation to increase by 

\[
p(d,a^*) + (-p_d/p_{1a})p_1(a^*) > 0.
\]

(This follows from applying the envelope theorem to the derivative of 

\[
p(d,a^*)S + a^* + p_1(a^*)S_1(S) \text{ with respect to } S,
\]

when \( S_1 \) is treated as a function of \( S \) whose derivative is 

\[
\frac{dS_1}{dS} = -\frac{p_d}{p_{1a}}.
\]

---

38 Note that, as discussed in Section 3.2.2.2., the increase in \( S_1 \) also adds to underlying deterrence.
39 See notes 14 and 15.
40 That is, 

\[
-p_d(S_{aa} + S_1p_{1a}) + (-p_d/p_{1a})(-p_{1a}(S_{aa} + S_1p_{1a})) = 0.
\]
Given that \( S \) and \( S_1 \) are fines, such increase in the expected effective sanction does not incur additional sanctioning costs: \( c(S) = c(S_1) = 0 \) for all \( S \) and all \( S_1 \). Therefore, the coordinated increase in the two fines can play the role played by increasing the fine for the underlying violation in Becker’s (detection avoidance-free) argument for high fines and low detection probabilities.

Thus, the fact that existing models leave out the possibility that detection avoidance may itself be sanctioned is decisive for the literature’s main results. It must be reiterated, however, that this analysis assumes that detection of detection avoidance is not itself avoidable, a possibility discussed in the next section.

6 The Recursivity of Detection Avoidance

One might be tempted to regard detection avoidance as just another hidden action to which the vast literature on incentives\(^{42}\) – inclusive of the conventional enforcement model – may be brought to bear. But a fundamental and differentiating characteristic of detection avoidance is that it is a recursive activity. Thus, in the context of the conventional enforcement model, sanctioning any activity – including detection avoidance – generates additional effort to avoid detection of that activity. The recursivity of detection avoidance thus spins out a potentially infinite sequence of ever greater orders of detection avoidance. Sanctioning the underlying offense encourages “first-order” detection avoidance. Sanctioning first-order detection avoidance encourages “second-order” detection avoidance. Sanctioning second-order detection avoidance encourages third-order. Sanctioning third encourages fourth. And so on.

The recursivity of detection avoidance was first pointed out and analyzed in Sanchirico (2006a, 2006b). This section of the chapter first discusses the practical import and conceptual nature of detection avoidance’s recursivity. It then considers how such recursion effects the efficacy of sanctioning detection avoidance.

6.1 The Practical Import of Detection Avoidance’s Recursivity

Is it plausible that violators engage in, say, detection avoidance four times removed? Arguably, it is conditionally plausible and that is all it needs to be. Fourth-order detection avoidance is plausible, that is, if it is assumed that the government can identify, sanction, and thereby discourage

\(^{41}\) This section is drawn from Sanchirico (2006a) and the corresponding formal analysis in Sanchirico (2006b).

\(^{42}\) Kreps (1990), chapter 16, describes this literature.
detection avoidance three times removed – which is the only case in which the plausibility of fourth-order detection avoidance matters. The effectiveness of the sanction on third-order detection avoidance presupposes that third-order detection avoidance is a discernible activity in the minds of violators. And, at that point, the pedigree of the activity becomes irrelevant: it is just an activity and it will be covered up like any other that is also subject to sanction.

It is thus important to keep in mind that the real impact of detection avoidance’s recursivity lies not in the full conceptual stretch to infinity. It lies, rather, in the fact that detection avoidance always grows another head, and so remains one order greater than the last effective order of sanction.

6.2 The Nature of Detection Avoidance’s Recursivity

It is also important to distinguish the recursivity of detection avoidance from the well-recognized policy pitfall of ignoring substitution effects: as when limiting boat size induces fishermen to use better equipment with little effect in reducing harvest; or as when rewarding for high student test scores quashes the teaching of unobservable attributes like creativity. In these cases, a single activity is effectively taxed (respectively, subsidized), and the corresponding reduction (respectively, increase) in that activity makes an alternative activity more (respectively, less) productive. The interactive mechanism is a “cross effect” in the violator’s optimization problem – a substitution effect to be precise.

The recursivity of detection avoidance does not describe a situation in which lower-order detection avoidance is taxed and thereby causes a substitution into higher. (Indeed, as discussed below, orders of detection avoidance are likely to be complements rather than substitutes.) Rather, what is happening is that the tax on lower-order detection avoidance simultaneously acts as a subsidy on higher-order avoidance. An additional dollar of sanction on \( n - 1 \)th-order avoidance is in effect an additional dollar of reward for \( n \)th. Thus the interactive mechanism across orders of detection avoidance comes not from cross-effects in the violator’s effective expected sanction minimization problem, but rather from the fact that the policy instrument – the sanction on \( n - 1 \)th-order avoidance – has two points of impact on that minimization problem.

Formulaically, allowing for infinite orders of detection avoidance and sanctions thereon, the effective expected sanction for the underlying

---

43 See, for example, Holmstrom and Milgrom (1991).
44 The technical material that follows is derived from Sanchirico (2006b).
violation is the minimized value of \( (p_0 S_0 + a_1) + (p_1 S_1 + a_2) + (p_2 S_2 + a_3) \) 
. . . , or more precisely,

\[
\min_{d_0, a_1, a_2, \ldots} \left[ p_0 \left( \frac{d_0}{S_0} \right) + \sum_{i=1}^{\infty} p_i \left( \frac{d_i}{S_i} \right) \right]
\]

In this expression, \( a_i \) is the potential violator’s \( i \)th-order detection avoidance expenditure, \( p_0 \) is the probability that the underlying violation is detected, \( p_i \) is the probability that the violator’s \( i \)th-order detection avoidance is detected, \( d_0 \) is public expenditure on detection of the underlying violation, \( d_i \) is public expenditure on detection of \( i \)th-order detection avoidance, \( S_0 \) is the ex post sanction on the underlying activity, and \( S_i \) is the ex post sanction on \( i \)th-order detection avoidance.

The pluses and minuses in the expression show the assumed signs of first derivatives.\(^{45}\) For example, the probability \( p_0 \) that the underlying violation is detected increases in public detection effort \( d_0 \) and decreases in first-order detection avoidance effort, as well as higher-order detection avoidance efforts. For another example, the probability \( p_2 \) that second-order detection avoidance is detected increases in second-order detection avoidance itself \( a_2 \) and in detection effort \( d_2 \) directed at second-order detection avoidance, and decreases in third-order detection avoidance \( a_3 \) as well as higher orders of detection avoidance.

The contradictory effect of sanctioning a given order of detection avoidance can be seen from this expression. Consider the addend in the middle summation for some arbitrary counting order \( i' \geq 1 \). Sanctioning \( i' \)th-order detection avoidance “taxes” \( i' \)th-order detection avoidance – according to the amount that such \( i' \)th-order detection avoidance increases the chance that \( i' \)th-order detection avoidance itself is detected. But sanctioning \( i' \)th-order detection avoidance also “subsidizes” all orders of detection avoidance that are higher than \( i' \) – according to the amount that such higher orders of detection avoidance decrease the chance that \( i' \)th-order detection avoidance is detected.

What effect does the tax/subsidy which is \( S_i \) have on various orders

\(^{45}\) I shall make assumptions regarding all \( p_i \) that are analogous to those made for \( p \) and \( p_1 \) in notes 14 and 15. In particular, I shall assume that \( p_{i-k} > 0 \) for all positive \( k \) such that \( i - k \) is not less than zero. Thus, higher-order detection avoidance decreases the probability of all lower-order detection at a decreasing rate.
of detection avoidance? The answer is potentially complicated by cross-effects of the form \( p_{i_{\alpha}} \). Because of such cross-effects, one cannot say whether detection avoidance of any given order will increase or decrease as a result of increasing the sanction of any given order.\(^{46}\) Let us, therefore, assume that such cross-effects are negligible. (For a further discussion of cross-effects see Section 6.4.) Then the left-hand side of the first-order condition for any given order \( i \) of detection avoidance, which is

\[
\frac{\delta a_i}{\delta s_i} = p_{i_{\alpha}} S_i + \sum_{j=0}^{i-1} p_{j_{\beta}} S_j + 1 = 0
\]

is independent of the level of all other orders \( k \neq i \) of detection avoidance. The implicit function theorem may, therefore, be individually applied to each order of detection avoidance. And it may thereby be determined that the effect on \( i \)th-order detection avoidance of raising the \( i \)th-order sanction itself is negative:

\[
\frac{\delta a_i}{\delta s_i} = -p_{i_{\alpha}} \left( p_{i_{\alpha}} S_i + \sum_{j=0}^{i-1} p_{j_{\beta}} S_j \right) < 0,
\]

while the effect on \( i \)th detection avoidance of raising a sanction of order less than \( i \) is positive: for all \( i > k \geq 0 \),

\[
\frac{\delta a_i}{\delta s_k} = -p_{i_{\alpha}} \left( p_{i_{\alpha}} S_i + \sum_{j=0}^{i-1} p_{j_{\beta}} S_j \right) > 0.
\]

To answer the initial question regarding the impact of increasing a particular \( S_r \) on various orders of detection avoidance, we can turn the foregoing analysis around and fix the sanctioning order rather than the order of detection avoidance. Assuming small cross-effects, increasing the \( i \)th-order sanction \( S_i \) decreases \( i \)th-order detection avoidance, but increases all higher orders of detection avoidance \( i > i' \).

6.3 Sanctioning Hierarchies

It was shown in Section 5.2 that adding sanctions for first-order detection avoidance can restore Becker’s maximal fine prescription when higher orders of detection avoidance are ignored. More generally, it was shown that the simultaneous sanctioning of underlying violations and detection

\(^{46}\) The ambiguity caused by cross-effects is a common phenomenon in economic theory.
avoidance can counteract the positive effect on detection avoidance of increasing the sanction for the underlying violation. The question then arises: does the same idea carry over to a model that acknowledges the recursivity of detection avoidance?

It is true that, in such a model, there are more orders of detection avoidance to be counteracted. But it is also true that there are also more orders of sanction to be applied. Thus, although there is always yet another order of detection avoidance no matter how far out sanctions extend, there is always another higher order of sanction to be imposed. Thus, perhaps sanctioning all orders simultaneously would have an effect analogous to that described for the single order case in Section 6.2.

Sanchirico (2006a, 2006b) studies this question. He begins with the point that uniformly sanctioning all orders of detection avoidance – that is, imposing the same sanction on all orders of detection avoidance – is likely to increase, not decrease, detection avoidance – and at every order. He then shows that the kind of variable sanctioning structure that would be required to reduce detection avoidance is ever increasing across orders of detection avoidance, and so, he concludes, impractical. These findings are reviewed in the remainder of this section.

6.3.1 Uniform sanctions hierarchies Consider, first, a uniform sanction for all orders of detection avoidance as well as the primary violation: that is, for all \( i = 0,1,2,3, \ldots S_i = S > 0 \). Such a sanctioning structure resembles several aspects of the law as written\(^{47}\) (if not all aspects, and if not as such law-enforced): all orders of perjury, including perjury about perjury about perjury about perjury, are potentially perjury and are sanctioned, in theory, to the same degree. The same holds for all orders of obstruction of justice.

The first-order condition for choice of \( i \)th-order detection avoidance, (7.1), simplifies to

\[
\left( p_{m1} + \sum_{j=0}^{i-1} p_{mj} \right) S + 1 = 0.
\]

The first addend is the marginal net increase, due to \( i \)th-order detection avoidance, of the effective expected sanction. The second addend, 1, is the marginal cost of \( i \)th-order detection avoidance. Given fulfillment of the first-order condition, the first addend must be negative, which is to

\(^{47}\) Section 2.1 describes laws regulating evidentiary foul play. See also Sanchirico (2004a).
say that the violator must be choosing \( i \)th-order detection avoidance at a level such that marginal \( i \)th-order detection avoidance further reduces the effective expected sanction. More specifically, marginal \( i \)th-order detection avoidance must reduce the expected sanction per dollar of \( S \):

\[
p_{ia_i} + \sum_{j=0}^{i-1} p_{ja_j} < 0. \tag{7.2}
\]

It follows that, even though \( S \), the sanction that applies to all orders of detection avoidance, is both a sanction and reward for \( i \)th-order detection avoidance, raising \( S \) raises, on net, the marginal benefit to the violator of \( i \)th-order detection avoidance. This, in turn, implies that \( S \) raises the level of \( i \)th-order detection avoidance.

Formally, applying the implicit function theorem to the first-order condition (continuing to assume that cross-effects are negligible) and using (7.2) to sign the numerator, we have:

\[
\frac{\partial a_i}{\partial S} = -\frac{p_{ia_i} + \sum_{j=0}^{i-1} p_{ja_j}}{p_{ia_i} + \sum_{j=0}^{i-1} p_{ja_j}} > 0.
\]

Note that the sign of this derivative is positive for all orders \( i \) of detection avoidance. Therefore, all orders of detection avoidance increase as a result of increasing the uniform sanction.\(^{48}\)

As an aside, consider, by contrast, the effect on the underlying violation of increasing the uniform sanction. By the envelope theorem, the effective expected sanction increases by the sum of the detection probabilities, \( p_0() + \sum_{i=1} \pi_i p_i > 0 \). Thus, although all orders of detection avoidance increase, fewer underlying violations occur. Why the contrary effect on the underlying violation? The underlying violation is different than all orders of detection avoidance. In the case of the underlying violation, and only in the case of the underlying violation, there is no lower-order sanction that the underlying violation helps to avoid. Hence, there is no lower-order sanction whose increase acts to encourage more of the underlying violation.

6.3.2 Variable sanctioning hierarchies

So far, we have considered only uniform sanctioning hierarchies, wherein \( S_i = S \), for all \( i \). And we have

\(^{48}\) Sanchirico (2006a) provides a more detailed description of the intuition behind this result.
seen that increasing the uniform sanction $S$ may well encourage all levels of detection avoidance. In theory, however, the government can adjust the sanctioning hierarchy so as to increase the effective expected sanction for the underlying violation without also increasing detection avoidance of order $i$. To do so, it must raise the sanction for order $i$ at a faster rate than it raises the sanction on lower orders of detection avoidance (including the primary violation as order 0). More precisely, and as shown below, the rate of increase in ever higher-order sanctions must be greater than a particular weighted average of the rate of increase at lower orders of sanction.

Imposing an increasing sanctions hierarchy corresponds to the fix described in Section 5.2 for the case of a single order of detection avoidance. However, as discussed below, imposing an ever increasing sanctions hierarchy is unlikely to be feasible in practice.

6.3.2.1 THE THEORETICAL POSSIBILITY OF PREVENTING INCREASES IN SOME OR ALL ORDERS OF DETECTION AVOIDANCE

Given numbers $dS_i$ for all $i$, write

$$S_i(\beta) = S_i + \beta dS_i,$$

for all $i$. Changing the parameter $\beta$ changes the sanctioning hierarchy in a particular direction defined by the given sequence of $dS_i$. The violator’s first-order condition for $i$th-order detection avoidance, (7.1), can then be rewritten as

$$\frac{d}{d\beta} p_{i,a} S_i(\beta) + \sum_{j=0}^{i-1} \frac{d}{d\beta} p_{j,a} S_j(\beta) + 1 = 0. \quad (7.3)$$

Continuing to assume negligible cross-effects, the implicit function theorem applied to the parameter $\beta$ then implies

$$\frac{\delta a_i}{\delta \beta} = -\left(p_{i,a} dS_i + \sum_{j=0}^{i-1} p_{j,a} dS_j\right) \left(\frac{\sum_{j=0}^{i-1} p_{j,a} S_j}{\sum_{j=0}^{i-1} p_{j,a} S_j} \right),$$

where this derivative is evaluated at $\beta = 0$.

Our question is this: for which sequences of $dS_i$ is the foregoing derivative non-positive. Assuming, as we shall throughout, that $S_k > 0$ for all $k$, this is the same as asking: for which sequences is the term in parentheses in the numerator non-negative,
Detection avoidance and enforcement theory

\[ p_{ia_i} dS_i + \sum_{j=0}^{i-1} p_{ja_i} dS_j \geq 0 \Leftrightarrow p_{ia_i} dS_i \geq - \sum_{j=0}^{i-1} p_{ja_i} dS_j. \]  

(7.4)

Now, from the first-order condition, (7.3), we have

\[ p_{ia_i} S_i + 1 = \sum_{j=0}^{i-1} \frac{p_{ja_i}}{S_j} S_j. \]

Both sides of this equality are positive. Dividing both sides of (7.4) by the respective side of such equality yields:

\[ \left( \frac{p_{ia_i}}{p_{ia_i} S_i + 1} \right) \frac{dS_i}{S_i} \geq \sum_{j=0}^{i-1} \frac{p_{ja_i} S_j}{\sum_{j=0}^{i-1} p_{ja_i} S_j} \frac{dS_j}{S_j}. \]

The right hand-side of this inequality is a weighted average (with positive weights adding to one) of the \( dS_j / S_j \) from \( j = 0 \) to \( j = i - 1 \). These are the percentage increases in the sanction at orders of detection avoidance below \( i \) (including the underlying violation). Each such lower order is weighted by its contribution to the marginal reduction in the effective expected sanction caused by increasing \( a_i \). This weight represents the relative importance of such lower order in the benefit to the violator of increasing \( i \)th-order detection avoidance, that benefit being the reduction in lower-order expected sanctions.

Thus, writing \( \alpha_j \) for the importance-weight corresponding to lower-order \( j \), the condition is fulfilled if and only if

\[ \frac{dS_j}{S_j} \geq \left( 1 + \frac{1}{p_{ia_i} S_i} \right) \sum_{j=0}^{i-1} \frac{\alpha_j}{S_j} \frac{dS_j}{S_j}. \]  

(7.5)

This condition implies that in order to prevent the sequence \( dS_j \) of sanction changes from causing an increase in \( i \)th-order detection avoidance it is necessary that the percentage increase in the sanction at \( i \) exceed or equal the importance-weighted average of the percentage increases in the sanction at all orders lower than \( i \). Furthermore, it is sufficient for preventing an increase in \( i \)th-order detection avoidance that the percentage change in \( S_j \) be \( 100 \times 1/(p_{ia_i} S_i) \) percent greater than such weighted average of lower-order percentage changes.

Extrapolating, in order for the sequence of sanction changes not to cause an increase in any orders of detection avoidance \( i \geq 1 \), it is necessary and sufficient that condition (7.5) hold for all \( i \geq 1 \).
Consider, for example, the special case in which \( a_i \) affects only \( p_{i-1} \), the
probability of detection of the adjacent lower order, and not the detection probability at other even lower orders such as \( i - 2 \), \( i - 3 \) etc. Then condition (7.5) reduces to

\[
\frac{dS_i}{S_i} \geq \left( 1 + \frac{1}{p_{i-1}S_i} \right) \sum_{j=0}^{i-1} \alpha_j \frac{dS_j}{S_j} \Leftrightarrow \frac{dS_i}{S_i} \geq \left( 1 + \frac{1}{p_{i-1}S_i} \right) \frac{dS_{i-1}}{S_{i-1}}
\]

Thus, the percentage increase in the sanction must be

\[
100 \times \frac{1}{p_{i-1}S_i}
\]

percent greater at each successively greater order of detection avoidance.

Suppose, for example, that the initial sanctioning structure is uniform across orders at the value \( S = 100 \). Further, suppose that for all \( i \), the marginal impact of \( i \)th-order detection avoidance on its detection probability is \( p_{i-1} = 0.10 \). Then, in order to avoid increasing detection avoidance at any order, the percentage increase in the sanction at order \( i \) (which equals the absolute increase given \( S = 1 \)) must grow at a rate of 10% in \( i \). It follows that the increase in the sanction on eighth-order detection avoidance must be twice that for the underlying sanction, the increase in 25th-order detection avoidance must be ten times that for the underlying sanction, and so on.

6.3.2.2 PRACTICAL ASSESSMENT Imposing ever increasing sanctions across higher and higher orders of detection avoidance supposes that the state can reliably distinguish between such orders. This seems unlikely, especially given the incentive that is thereby created for perpetrators to portray their higher-order avoidance activities as being of lower order – yet another form of detection avoidance. (Note that the problem here is not an inability on the part of the government to observe detection avoidance activities. Rather the problem is the government’s inability to determine whether a particular instance of detection avoidance that it observes is of order \( i \) rather than order \( j \).)

Indeed, there is some reason to believe that the best the state can do as a practical matter is a decreasing hierarchy of sanctions. Arguably, higher orders of detection avoidance are more likely to get lost in the crowd of daily activity. Indeed, in US Federal law, despite the fact that obstruction of justice of any order is technically obstruction of justice, higher orders of obstruction are, in fact, unlikely to result in separate charges, and are more likely to be punished with sentencing enhancements for the lower-order activity. These sentencing enhancements produce in effect a decreasing sanctions hierarchy. By one reading of the Federal sentencing
guidelines, for instance, first-order obstruction is punished by 18 months in prison, second-order by six, and third-, fourth-, fifth-, etc. by zero.\textsuperscript{49}

Therefore, the practical prospects for controlling detection avoidance by means of sanctioning – even sanctioning that in principle extends to all orders of detection – seem dim. A somewhat weaker but still decisive claim, considered in Section 7, is that sanctioning detection avoidance is a relatively \textit{inefficacious} way of controlling it, given the possibility of alternative “technological” approaches.

It must be reiterated, however, that such conclusions have been derived under the assumption that cross-effects are small. Let us now turn to a discussion of cross-effects.

\subsection*{6.4 Accounting for Complementarities Across Orders of Detection Avoidance}

In previous subsections it was assumed that there were no cross-effects across orders of detection avoidance. As Sanchirico (2006a) notes,\textsuperscript{50} however, higher and lower orders of detection avoidance are likely to be “complements.” More of either increases the productivity of the other. More cover-up of the cover-up, that is, makes the underlying cover-up itself more productive. Conversely, more cover-up makes covering up the cover-up more productive.

The fact that orders of detection avoidance are complementary in the sense just described complicates, but does not fundamentally alter, the conclusion that sanctioning detection avoidance is relatively inefficacious.

Taking account of complementarities across orders of detection avoidance, consider sanctioning \textit{only} first-order detection avoidance. This has two countervailing effects on second-order detection avoidance. First, as already discussed, the sanction on first-order detection avoidance acts as a reward for second-order detection avoidance. Second, because sanctioning first-order detection avoidance discourages first-order detection avoidance itself, and because first- and second-order detection avoidance are complementary, sanctioning first-order detection avoidance will, through this second channel, act to discourage second-order detection avoidance.

Thus, the direct effect of the first-order sanction – operating through the first-order sanction’s other role as a second-order reward – is to encourage second-order cover-up, while the indirect effect of the first-order

\textsuperscript{49} For details, see Sanchirico (2004a, 2006a).

\textsuperscript{50} Following Sanchirico (2006a, 2006b), Nussim and Tabbach (2009) also make reference to complementarity. Nussim and Tabbach only consider such complementarity between the primary violation (zero-order detection avoidance) and first-order detection avoidance.
sanction – operating through complementarities – is to discourage second-order cover-up. Either effect may predominate. Thus, incorporating complementarities, it is not possible to say whether sanctioning first-order cover-up will increase or decrease second-order cover-up.

However, any measure that reduces first-order detection avoidance – whether that measure be a sanction on first-order cover-up or a technological restructuring of evidentiary process, as discussed in Section 7 immediately below – benefits from the indirect reduction in second-order detection avoidance that operates through complementarity. The difference between the technological approach, described in the next section, and the sanctioning approach is that the technological approach gives this benefit free reign, while the sanctioning approach hampers it by simultaneously subsidizing second-order detection avoidance, as has just been described.

7 Technological Approaches to Detection Avoidance
As pointed out in Section 3.2.2.2.2, sanctioning is not the state’s only means of dealing with detection avoidance. Given the state’s choice of detection activity \(d\), the government effectively presents to the violator not a single probability of detection, \(p(d)\), but a “menu” of detection probabilities, \(p(d, a)\) one for each of her possible choices \(a\) of detection avoidance intensity. By adjusting the shape of this menu (by the means described below), the state can affect both violators’ detection avoidance choices and their decision whether to violate the law in the first place.

Sanchirico (2006a, 2006b) points out the existence of this alternative policy instrument and further argues that such a “technological approach” to detection avoidance may be a more efficient and effective means of reducing detection avoidance than sanctioning.

7.1 The Mechanics of the Technological Approach
A simple diagram from Sanchirico (2006b) helps to clarify the nature and effect of technological approaches to detection avoidance. Thus, suppose that we wish to reduce detection avoidance and imagine, for simplicity, that there are no pre-existing sanctions on detection avoidance.

The horizontal axis in Figure 7.1 depicts (first-order) detection avoidance effort \(a\) (in dollars). The vertical axis depicts the expected sanction per se for the underlying violation, the product of the probability of detection for the underlying violation \(p(a)\) and the sanction for the underlying violation \(S\). (Note that this “expected sanction per se” does not account for private detection avoidance costs.) The downward sloping line(s) (focus on the lower one for now) shows the inverse relationship between detection avoidance spending \(a\) and the expected sanction per se: the more the individual spends on detection avoidance, the lower the
Detection avoidance and enforcement theory

This curve can also be thought of as the violator’s “production function” for detection avoidance. Put another way, it describes the “technology” of detection avoidance in the same way that a firm’s production function describes its technology of production. In the technology of avoidance, detection avoidance is the costly input, and a reduced expected sanction per se is the output.

Notice that detection avoidance is assumed to reduce the expected sanction at a decreasing rate (an assumption that I have imposed throughout). This reflects the assumption that detection avoiders employ the most productive avoidance measures first and must reach deeper down into the barrel (or higher up into the fruit tree) as they expend more and more effort and expense on avoidance.

As in the analysis in Section 3.2.2.1 above, the violator, who wishes to minimize \( p(a)S + a \), the sum of the expected sanction per se and the cost of the avoidance, will choose \( a \) so that the marginal reduction in the expected sanction \(-p_a(a)S\) equals the marginal cost of an additional dollar of detection avoidance spending, which is one dollar by definition. Therefore, she will choose her detection avoidance activity at a point where the slope of

\[ \frac{d}{da}(p(a)S + a) = 0 \]

See, for example, note 14.

---

Figure 7.1  A technological approach to detection avoidance
the curve \( p(a)S \) equals \(-1\). For the lower curve, this is the point corresponding to \( a^0 \) on the horizontal axis and \( p^0 \) on the vertical.

A technological approach to reducing detection avoidance shifts the technology of detection avoidance. Specifically, ideally implemented, it shifts the expected sanction per se curve according to two requirements: (1) it uniformly reduces the marginal productivity of detection avoidance activity; and (2) it does not lower the probability of detection in the absence of avoidance activity (i.e. at \( a = 0 \)). It therefore flattens the curve describing the relationship between avoidance and expected sanction without lowering (and possibly even raising) its level at \( a = 0 \). Such “upward flattening” is depicted by movement from the lower curve to the upper curve in the figure.

This upward flattening has two effects. First, violators spend less on detection avoidance. The slope of the upper curve flattens to a slope of \(-1\) at a lower level of detection avoidance expenditure. Second, the probability of detection of the underlying activity is greater. This implies that deterrence of the underlying activity has increased.\(^5\) (As in Section 3.2.2.1 above, the envelope theorem dictates that marginal adjustments to detection avoidance activity \( a \) have no effect on deterrence.)

Therefore, the technological approach decreases detection avoidance costs while increasing deterrence of the underlying activity. (As discussed in Section 7.2 below, it may also increase the public costs of detection.)

It is important to recognize that the technological approach is not merely a matter of “making detection avoidance harder.” That is, requirement (1) above is not merely a matter of increasing the cost to the violator of reducing the probability of detection by an additional percentage point. Doing just this may lead violators to try harder, and thereby to increase, rather than decrease, the expenditure of resources on detection avoidance. Rather, the first requirement of the technological approach requires something more: that the productivity of avoidance effort and expenditure be reduced. To reduce the productivity of detection avoidance is to reduce the number of percentage points by which an additional dollar of detection avoidance reduces the probability of detection. As Sanchirico (2006a, 2006b) shows, this second requirement implies, but is not implied by, an increase in the cost of reducing the probability of detection.

\(^5\) Following Sanchirico (2006a, 2006b), Nussim and Tabbach (2009) identify the same two effects.
7.2 Public Detection Costs and the Comparison to Sanctioning
Reducing the productivity of detection avoidance is likely to come at the price of additional public costs. Decreasing the effectiveness of the detection avoidance dollar may, for example, require more costly surveillance systems or more lengthy and numerous interrogation sessions.

Yet if the issue is relative efficacy, such public costs cannot be considered in isolation, but must be compared with the public costs of the sanctioning approach. The sanctioning approach to detection avoidance requires additional, costly processes, which may be quite substantial. Perjury and obstruction also must be investigated and prosecuted.

Therefore, the sanctioning approach and the technological approach to detection avoidance are similar along two dimensions in the social calculus, but quite different along a third. They both increase primary activity deterrence and they both incur public costs. But because the technological approach is not prone to the recursivity that plagues sanctioning, the technological attack is more effective at reducing private detection avoidance costs.

7.3 Implementing the Technological Approach
How can the marginal productivity of detection avoidance be reduced (while also not increasing its total product)?

Merely devoting additional public resources to detecting violations will not suffice. Simply questioning yet another witness, for example, will not necessarily decrease the productivity of detection avoidance spending. If, without coaching, a witness’s answers will increase the violator’s chance of having to pay a $100,000 sanction by ten percentage points, but with $5000 of “preparation” this can be wholly prevented, then interrogating an additional witness will most likely increase, rather than decrease, the productivity of detection avoidance spending. Rather, to implement the technological approach, public detection spending must be specifically channeled so that each dollar and each unit of effort spent avoiding detection buys less of a reduction in the probability of detection. This is essentially a matter of making detection avoidance more difficult at each step – so that, for example, $5000 of witness coaching only partially prevents the witness’s positive impact on the probability of detection.

Sanchirico (2006a) suggests accomplishing this by designing evidentiary process so as to exploit and amplify the difficulties generally encountered in all human endeavors. Two difficulties – of cognition and of cooperation – are already exploited by current evidentiary process and may hold further potential.

An earlier paper, Sanchirico (2004b), reviews in detail how the law
Procedural law and economics

of evidentiary procedure exploits cognitive limitations. Imagine, for example, that the offender wishes to reduce the probability of detection by supplying a witness to swear falsely that the offender did not commit the underlying crime. Exploiting the witness’s cognitive limitations, the law takes several steps to reduce the productivity of time and effort spent preparing this witness. Consider three specific aspects of how testimony, depositions, and interrogations generally proceed. First, the witness will usually not see the questions in advance. Time spent preparing answers to the questions that one can anticipate is thereby less productive for the fact that such preparation may well be rendered much less valuable with a few poorly improvised answers to questions that were unexpected. Second, the questioner need not commit to her questions ahead of time, but may rather adjust the subject or tenor of additional questions based on what she perceives to be uncertainties and inconsistencies in the answers provided to previous questions. This renders preparation less productive because the witness is largely denied the opportunity of playing the odds that particular topics will not be “tested.” The test is not written ahead of time. If it becomes apparent that the witness is less prepared for a particular line of questioning, the questioner may shift course and emphasize that line of questioning. Conversely, if it becomes apparent that the witness has prepared for a given line of questioning, the questioner can choose to move on to other topics. Preparation time is thus generally rendered less productive. Lastly, interrogations and depositions exploit the very real effects of fatigue. The difficult task of fabricating testimony becomes all the more difficult as the fabricator tires. While interrogators and deposers may substitute in and out during questioning, the witness is on her own. Hours of preparation can be rendered virtually ineffective by a few unguarded answers in the last few moments of a long day of questioning.

Game theorists have long recognized the possibility of exploiting the difficulties and fragilities of coordination and cooperation among multiple agents. These lessons apply to the state’s efforts to reduce the productivity of detection avoidance activity. Detection avoidance, like any human activity, often requires or is facilitated by coordination among several individuals, especially if it is effected on a large scale. The state can play these individuals against each other by structuring interrogation and prosecution to amplify the temptation to break ranks. Specific practical techniques employed by law enforcement in this area include, first, the hearsay exception for statements of a co-conspirator. Statements made by a co-conspirator (during the pendency of the conspiracy and in furtherance thereof) may be used substantively against a party even if they are not made for the purpose of testifying in the
current case. Other devices include prosecutorial immunity, plea agreements, no prosecution agreements, special protection for whistleblowers, and rewards for informants. All of these make cooperation in detection avoidance harder to maintain and thus reduce the usual productivity gains from teamwork.

8 Conclusion
The state’s efforts to detect violations play a leading role in the conventional model of economic enforcement. Violators’ efforts to avoid detection, on the other hand, are barely alluded to. Yet the best data available suggest that, in reality, detection and detection avoidance share the stage – that they are indeed yin to each other’s yang. Law and economics has just begun to bridge this important gap between the way it understands enforcement and the way that enforcement actually proceeds. More work remains to be done – especially with regard to policy tools, such as sanctions or technological adjustments to legal process, that are directed at detection avoidance itself.

Bibliography
Procedural law and economics


Detection avoidance and enforcement theory