The Conspirator Dilemma: 
Introducing the “Trojan Horse” Enforcement Strategy

Omri Yadlin*

ABSTRACT

The Trojan Horse enforcement mechanism turns law-breakers into law-enforcers by entrusting them with the right to file a private suit and to collect a fine from their accomplices. Unlike leniency or state witness programs, the Trojan Horse mechanism is not dependent on an effective public enforcement agency operating in the background. It positions conspirators in a prisoner-dilemma at the very first stage of their conspiracy, long before the public enforcers are on their tails. Thus, the Trojan Horse mechanism enables the state to decriminalize certain conspiracies.

As shown in this paper, the Trojan Horse mechanism has a (comparative) advantage in asymmetric settings, such as in the employment of illegal immigrants. Whereas fines and other criminal sanctions hardly deter empty-pocket lawbreakers like illegal immigrants, the carrot this mechanism waves might lure illegal immigrants into suing their employers and leaving the country with a prize in their pockets. Thus, it provides a more humane enforcement mechanism than the prevailing ones and creates a strong deterrence effect on employers who consider hiring illegal immigrants.

*Tel Aviv University, Law School. This paper benefited from discussions with Malcolm Feeley, Ariel Porat, Alon Harel, Dan Rubinfeld, Jesse Fried, Omri Ben-Shahar, Sanford Kadish, Ehud Kamar, Eric Talley, and Ron Shapira. While developing the paper, I learned that Bob Cooter and Nuno Garupa were independently developing a similar idea with a different emphasis. While my paper addresses conspiracies in general, Cooter
I. INTRODUCTION
Scholars have long debated the optimal division of labor between public and private law enforcement. Pioneers in this area of scholarship were Becker & Stigler, who called for the privatization of law enforcement. According to their proposal, private agents would collect fines from outlaws on a first-to-file-collects-the-fine basis. In response, Landes & Posner pointed at certain failures from which the market for private law enforcement suffers. They also questioned the novelty of Becker & Stigler’s proposal, noting that in practice, the law enforcement enterprise already occupies bounty hunters, class action representatives, and other legal entrepreneurs. On that basis, Landes & Posner concluded that there is no reason to...
question the efficiency of the current separation of responsibilities between the public and the private sectors of law enforcement.

This paper proposes a different kind of enforcement mechanism, one that does not rely on public agencies, victims or bounty hunters. The best enforcers of the law, goes my argument, are the most informed parties, namely, the law-breakers themselves. Thus, the mechanism I propose here entrusts the first felon filing a civil suit with the right to collect a portion of the fine from his accomplices; the rest of the fine will go to the government. Whereas public enforcers, in the course of investigating a case, may subject suspects to the prisoner dilemma by offering each one of them immunity for testifying against the others, the mechanism I propose here positions lawbreakers in the “conspirator dilemma” from the very moment they enter the illegal scheme. Each felon faces the risk that her partner might defect, and thus each partner’s best strategy will be to act as a *Trojan Horse* and collect the fine first. Hence, unlike public and private enforcement agencies, which typically start investigating a crime only after its completion, the *Trojan Horse* conspirator would sue his partners at the preliminary stages of the crime, often before any harm has been done.\(^3\) The mechanism I propose therefore

\(^3\)In a way, my point is motivated by the same vision that inspired Kaplow and Shavell’s paper “*Optimal Enforcement with Self-Reporting of Behavior,*” 102 J. Pol. Econ. 583 (1994), namely, that the law should be designed to induce wrongdoers to expose their own wrongdoing. However, the mechanisms differ in some very substantial ways. First, whereas self-reporting mechanisms are designed to deal with no-fault liability, the *Trojan Horse* mechanism addresses mainly intentional violations of the law. Second, self-reporting is an *ex post* mechanism in the sense that it induces law-breakers to report themselves after they realize that they have broken the law. The goal of the *Trojan Horse* mechanism, on the other hand, is to talk conspirators out of their plan *ex ante.*
decreases the level of trust among criminals and increases their costs of cooperation. Thus, this mechanism may serve as an effective deterrence apparatus against such conspiracies.

Clearly, the Trojan Horse mechanism can work alongside public enforcers. In such a competitive setting, in addition to collecting the fine, a felon who wins a Trojan Horse suit would be immune from government’s prosecution. However, one advantage of the Trojan Horse mechanism is that unlike leniency and state witness programs, it is not dependent on an effective public enforcement mechanism operating in the background, thus allowing us to decriminalize certain offences.

The following two examples demonstrate the way in which the Trojan Horse mechanism operates.

**EXAMPLE ONE: PRICE FIXING**
Consider first the case of price fixing arrangements. Currently, public agencies and private entities work in tandem to enforce the antitrust laws. But both of these enforcement mechanisms are seldom privy to information about the formation and execution of anti-competitive agreements. Hence, some price-fixing arrangements escape sanctions altogether; others are detected only long after initiation, often after the cartel breaks apart.

Compared with any public or private enforcer, cartel members possess far more information about the anti-competitive nature of their conduct. The first-felon-to-file-collects-the-fine rule takes advantage of this comparative advantage. Since each member of the cartel faces the risk of being sued by the other, the best strategy for each member must be to defect first and file a suit as soon as sufficient evidence to support the case is gathered. Hence, the Trojan Horse mechanism raises significantly the cost of entering such an anti-competitive agreement.

---

4The plaintiff’s immunity will be limited, of course, to those offences that the defendant was convicted for. To determine whether the conviction covers a certain criminal conduct or not, we may use criteria similar to those used under the double jeopardy doctrine.
EXAMPLE TWO: ILLEGAL EMPLOYMENT

Another good example is the import and employment of illegal immigrants. A large segment of the workforce in many industrial countries consists of illegal immigrants. Typically, the laws of these countries subject illegal immigrants and their employers to significant sanctions. The fact that these fines do not deter employers from hiring illegal immigrants suggests that public enforcement is ineffective. One explanation for this enforcement failure is that enforcement agencies are uninformed. Another explanation is that enforcement agencies are captured by interest groups that persuade the public enforcement agency to keep a high profile in the media, but a low presence in the sweatshops.

Arguably, private agencies, in line with Becker & Stigler’s proposal, are less influenced by interest groups and more informed than public agencies and, therefore, have the potential to somewhat tighten the enforcement of these laws. But the best enforcers of these laws are probably the illegal immigrants themselves, who know their status better than any public or private enforcer does. Under my suggestion, therefore, illegal immigrants will be authorized to collect fines from their employers and/or from the manpower agent who assigned them to their employer. If the immigrant wins his or her case, the court would order the defendant to deposit the fine at the country’s gate and the immigrant would collect it on his or her way out of the country.

Having introduced the mechanism, we can now turn to a discussion of its details. In the second part of this paper, below, I use a simple model to compare the operation of the Trojan Horse mechanism in a one-shot scheme to its functioning in n-shot and infinite, continuous schemes. The third part modifies the model to address asymmetric schemes like illegal employment.

II. THE TROJAN HORSE MECHANISM IN THREE SETTINGS

In this part, I examine the way in which the Trojan Horse mechanism operates in three different contexts. In the first scenario, two (or more) conspirators engage in only one incident of law-breaking. In the second, two conspirators engage in a series of $n$ transgressions. Finally, the third scenario examines the operation of the Trojan Horse mechanism in infinite, serial conspiracies.

The one factor I used to compare the efficacy of enforcement systems is the minimum fine required under each system to deter wrongdoing. Assuming risk neutrality and given the fact that the financial
resources of criminals are limited, criminals who cannot meet a certain fine will not be deterred by a mechanism whose effectiveness depends on such a fine. Thus, the lower the minimum fine, the more criminal activity the law deters.

A. The One-Shot Conspiracy Game

Consider the case of two agents, say, a homeowner name Roe and a contractor name Doe, who conspire to misreport their business to the tax authorities. Roe and Doe expect to derive from their illegal interaction a gain of $G$ each. The fine that the public enforcement agency imposes on parties to such a conspiracy is $F$, and the probability that the plot will be detected and successfully prosecuted is $P$. Assuming Roe and Doe are risk neutral, they would not engage in such a scheme if:

\[(1) \quad (1-P)G + P(G-F) < 0\]

The minimum fine that satisfies this condition is:

\[(2) \quad F^* = \frac{G}{P}\]

Now suppose we replace the public enforcement agency with Trojan Horses; namely, lawmakers rely solely on Roe and Doe to enforce the tax laws. Again, I assume that Roe and Doe gather sufficient evidence to support such a suit only after they have completed the tax evasion plot. At that time, the probability that a Trojan Horse’s suit will prevail is represented by $Q$ and the cost of filing such a suit by $C$. In the event that the Trojan Horse prevails at trial, the court orders the defendant to pay a fine ($F$) and assigns a portion of this fine ($R$, where $R \leq F$) to the plaintiff. The rest of the fine goes to the state.
The 2*2 table below presents the dilemma the two conspirators face immediately after they have executed their commitments under the agreement:

\[
\begin{array}{c|cc|c}
\text{Doe’s Payoff} & \text{Roe Cooperates} & \text{Roe Sues} \\
\hline
\text{Roe Cooperates} & G & Q(G+R)+(1-Q)(G)-C \\
\text{Roe Sues} & Q(G-F)+(1-Q)G & Q(G-F)+(1-Q)G \\
\end{array}
\]

Clearly, in the absence of public enforcement in the background, it is Pareto optimal for Roe and Doe to evade the tax laws and to refrain from suing one another. However, under the following condition, the Trojan Horse mechanism induces the two conspirators to race each other to the courthouse as soon as the illegal scheme has culminated:

\[(3) \quad Q(G+R)+(1-Q)G-C>G\]

Condition (3) can be reduced to \(QR>C\).

Knowing in advance -- i.e., before commencing the scheme -- that eventually the two conspirators will race each other to the courthouse, the two will engage in the scheme only if the following “disparticipation” condition is met:

\[(4) \quad \frac{Q(G-F)+(1-Q)G+Q(G+R)+(1-Q)(G)-C}{2}<0\]

Condition (4) can be reduced to \(Q(F-R)>2G-C\).

Solving for the minimum fine \((F^*)\) that would deter participation, we find that:

\[(5) \quad F^*=\frac{2G}{Q} \text{ and } R^*=\frac{C}{Q} \quad \text{if } C<2G\]

and
One conclusion we may derive from the above model is that the minimum fine, the fine that would persuade Roe and Doe out of their plan to evade the tax laws, is not negligible. In fact, we find that public enforcement and the *Trojan Horse* enforcement system will be equally effective only if the probability of the *Trojan Horse* prevailing at trial is twice the probability of the tax authorities detecting and successfully prosecuting the evasion. The intuition lying behind this result is quite simple: When the public enforcement mechanism is at play, the two conspirators are forced to pay the fine to the government. When, on the other hand, *Trojan Horses* are the only enforcers of the law, one conspirator is let off the hook and gets to keep his ill-gotten gains to himself -- the fine, therefore, should be higher.

Although this comparison seems to favor the public agencies, it does not suggest that *Trojan Horses* are inferior enforcers. Conspirators, by definition, always detect and, arguably, have superior access to evidence concerning the crime they commit. Thus, it might be reasonable to assume that the probability of them prevailing at trial ($Q$) is more than twice the success rate of the tax authorities ($P$), in which case *Trojan Horses* are more effective enforcers than public agencies.

Another consideration we should take into account is that the effectiveness of the *Trojan Horse* mechanism improves with the number of conspirators who participate in the illegal scheme. To demonstrate this point, assume exactly the same game, but with $N$ conspirators, where $N > 2$. Each agent faces the following dilemma: If all agents cooperate, namely, if they all refrain
from suing, each one of the accomplices will gain $G$. However, conspirators would rush to file a suit against their accomplices if:

(7) $R > C/Q$, where $R \leq (N-1)F$

To ensure that the $N$ agents do not conspire, their expected return must be negative:

(8) $[Q(G+R)+(1-Q)G-C] + [Q(G-F)+(1-Q)G](N-1)/N < 0$

Condition (8) can be reduced to $QF(N-1)-QR > NG-C$.

Solving for the minimum fine that would meet conditions (7) and (8), we find that:

(9) $F^* = NG/(N-1)Q$ and $R^* = C/Q$ for any $C < NG$

and

(10) $F^* = C/(N-1)Q$ and $R^* = C/Q$ for any $C \geq NG$

Based on the above conditions, we may conclude that the larger the number of conspirators, the more effective the Trojan Horse enforcement mechanism. In other words, with the Trojan Horse mechanism in the background, criminals would prefer to act in concert with a smaller number of players.

**B. THE FINITE N-SHOT CONSPIRACY**

Consider again the case of two conspirators, only now they embark on a series of $n$ incidents of lawbreaking -- $n$ tax evasions, $n$ bribes, or any other series of $n$ incidents of lawbreaking. Whereas in the one-shot game, the parties derive no benefit before the completion of the scheme, an $n$-shot scheme is divided into $n$ separate incidents, each incident enriching each

---

5 In addition, conditions (9) and (10) imply that as the number of conspirators increases, the reward allotted to the Trojan Horse should decrease, and portion of the fine that the government collects should increase.
conspirator by $g$. Another, more important distinction between $n$-shot and one-shot schemes is that whereas parties to a one-shot scheme can sue only after the scheme has concluded, each shot in the $n$-shot game provides conspirators with the evidence necessary to file a suit against their associates. Just as in the one-shot game, however, conspirators cannot sue before the first incident. For the purpose of comparison between the one-shot and the $n$-shot games, I assume the profits the parties derive from these schemes are equal, namely I assume $g=G/n$ (interest rate = zero). The cost of filing and the probability of winning the case remain $C$ and $Q$, respectively.

Looking at the positions of the two players after the $n$-th shot, they should appear very similar to their positions after the one-shot game. Hence, the two players will race to the court after conducting the last tax evasion if:

$$(14) \ R > C/Q$$

The disparticipation condition in the serial case, however, is different from that in the one-shot conspiracy. This time, we should not ask ourselves whether the players would embark on the entire series of tax evasions that provides them with a gain of $SG$, but, rather, whether they would embark on the last tax evasion, which yields only $g$. The following table presents their payoffs when they face the dilemma of whether to engage in the last tax evasion and sue thereafter or whether they should sue beforehand:

*Table B: The Finite n-Shot Game (at the n-1 Stage)*

<table>
<thead>
<tr>
<th>A’s Payoff</th>
<th>A Cooperates</th>
<th>A Sues</th>
</tr>
</thead>
<tbody>
<tr>
<td>B’s Payoff</td>
<td>A Cooperates</td>
<td>A Sues</td>
</tr>
<tr>
<td>A</td>
<td>$[Q(G-F)+(1-Q)(G+R)+(1-Q)(G-C)]/2$</td>
<td>$Q(G(n-1)/n-R)+(1-Q)(G(n-1)/n-C)$</td>
</tr>
<tr>
<td></td>
<td>$[Q(G-F)+(1-Q)(G+R)+(1-Q)(G-C)]/2$</td>
<td>$Q(G(n-1)/n-R)+(1-Q)G(n-1)/n$</td>
</tr>
<tr>
<td>B Cooperates</td>
<td>$Q(G(n-1)/n-F)+(1-Q)G(n-1)/n$</td>
<td>$[Q(G(n-1)/n-F)+(1-Q)G(n-1)/n+Q(G(n-1)/n+R)+(1-Q)(G(n-1)/n-C)]/2$</td>
</tr>
</tbody>
</table>
The disparticipation condition in the last stage of the game, therefore, is:

\[(15) \ Q(G(n-1)/n+R)+(1-Q)(G(n-1)/n)-C > \frac{[Q(G-F)+(1-Q)G+Q(G+R)+(1-Q)(G)-C]}{2}\]

This condition can be reduced to

\[FQ+RQ > 2g-C\]

Solving for the minimum fine that would meet these conditions, we find that:

\[(16) \ F^*=R^* = \frac{2(g-C)}{Q} \text{ if } C<2g\]

and

\[(17) \ F^*=R^* = \frac{C}{Q} \text{ if } C \geq 2g\]

If conditions (16) and (17) are satisfied, the parties will not engage in the \(n\)-th incident, and the parties will figure out that \(n-1\) is actually the last stage of their scheme. Thus, they will face the same dilemma at the \(n-2\) stage. By backward deduction, we reach the conclusion that under this set of conditions, the two parties would race to court immediately after the first incident.

Our goal, however, is to deter law-breaking all together, and thus, we should search for the condition under which the two parties would not embark on even the first incident. This is actually the one-shot game we presented earlier, only this time the gain from a lawbreaking incident is not \(G\), but \(g\). Solving for the minimum fine that would meet these conditions, we find that:

\[(18) \ F^* = \frac{2g}{Q} \text{ and } R^* = \frac{C}{Q} \text{ if } C<2g\]

and

\[(19) \ F^* = R^* = \frac{C}{Q} \text{ if } C \geq 2g\]

Clearly, these two conditions satisfy conditions (16) and (17) above, and thus they are binding.
C. INFINITE CONSPIRACY GAMES

In the finite games discussed above, the parties knew from the outset that the day would come when the scheme would end and both would face the existential dilemma of “to sue or not to sue.” In infinite schemes, on the other hand, conspirators face a different dilemma: to sue or to continue with the scheme. As shown below, this distinction has significant consequences.

For the purposes of simplicity, I will assume the following payoffs: At each stage of the crime, the parties derive a gain of $g$. To demonstrate, $g$ can be understood as the one-day gain each cartel member derives from price-fixing. Interest rate is $r$, and for consistency, I assume that $g/r = G$. Each incident of law breaking produces sufficient evidence to allow conspirators to prevail at trial with probability $Q$. Filing such a Trojan Horse claim costs $C$. In the event that the Trojan Horse plaintiff wins the case, the defendant pays a fine $F$ and the plaintiff collects a portion of this fine -- $R$. Thus, the payoff table the parties observe at each stage looks as follows:

<table>
<thead>
<tr>
<th>B’s Payoff</th>
<th>A’s Payoff</th>
<th>A Cooperates</th>
<th>A Sues</th>
</tr>
</thead>
<tbody>
<tr>
<td>B Cooperates</td>
<td>$g/r = G$</td>
<td>$g/r = G$</td>
<td>QR-C</td>
</tr>
<tr>
<td>B Sues</td>
<td>QR-C</td>
<td>-QF</td>
<td>[QR-C-QF]/2</td>
</tr>
</tbody>
</table>

In equilibrium, both parties would rush to court soon after the first violation if:

(20) $QR-C > G$
Assuming condition (20) is met, both conspirators know, before engaging in the first violation, that the next morning both will race to court. Hence, they would not engage in the first incident of law-breaking if:

\[(21) \frac{[QR-C-QF]}{2}+g<0\]

Equation (21) can be reduced to \(QF-QR>2g-C\).

Searching for the minimum fine that would satisfy these conditions, we find that:

\[(22) F^*=(\frac{G+2g}{Q}) \text{ and } R^*=(\frac{G+C}{Q}) \text{ if } C<2g\]

and

\[(23) F^*=R^*=(\frac{G+C}{Q}) \text{ if } C\geq 2g\]

**D. ANALYSIS**

The tables below summarize our findings.

**Summary Table for Minimum Fines (F*)**

<table>
<thead>
<tr>
<th></th>
<th>(C&lt;2g)</th>
<th>(2g\leq C&lt;2G)</th>
<th>(C\geq 2G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Shot Game</td>
<td>(\frac{2G}{Q})</td>
<td>(\frac{2G}{Q})</td>
<td>(\frac{C}{Q})</td>
</tr>
<tr>
<td>(n)-Shot Game</td>
<td>(\frac{2g}{Q})</td>
<td>(\frac{C}{Q})</td>
<td>(\frac{C}{Q})</td>
</tr>
<tr>
<td>Infinite Game</td>
<td>(\frac{(G+2g)}{Q})</td>
<td>(\frac{(G+C)}{Q})</td>
<td>(\frac{(G+C)}{Q})</td>
</tr>
</tbody>
</table>

**Summary Table for Minimum Rewards (R*)**

<table>
<thead>
<tr>
<th></th>
<th>(C&lt;2g)</th>
<th>(2g\leq C&lt;2G)</th>
<th>(C\geq 2G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Shot Game</td>
<td>(\frac{C}{Q})</td>
<td>(\frac{C}{Q})</td>
<td>(\frac{C}{Q})</td>
</tr>
<tr>
<td>(n)-Shot Game</td>
<td>(\frac{C}{Q})</td>
<td>(\frac{C}{Q})</td>
<td>(\frac{C}{Q})</td>
</tr>
</tbody>
</table>
D.1. The Effect of the Trojan Horse Mechanism on the Market for Crime

One intuitive conclusion that can be drawn from the first table is that the higher the filing costs, the less effective the Trojan Horse enforcement mechanism. Although this conclusion is applicable to any enforcement mechanism, it seems particularly important in our context. Filing a Trojan Horse suit is very costly for professional criminals who have a “respectable” past and a shining future in the outlaw community. Consider, for example, the case of a contractor who offers bribes to public officials. Filing a Trojan Horse suit would cut the stream of payments the contractor expects to receive in the future, because public officials would refrain from conspiring with him. Thus, in presence of the Trojan Horse mechanism, less agents will play in this market, but those who do play will be more professional.

Another implication is that in response to the introduction of the Trojan Horse mechanism, we should expect criminals to try to raise filing costs. For example, one route conspirators could be expected to take in response to the Trojan Horse threat is violence. Arguably, therefore, although the Trojan Horse mechanism would probably increase entry barriers to the market for crime and reduce the level of crime in society, at the same time, it might increase the level of violence among criminals.

D.2. The Effect of Divisibility and Infinity

Another interesting conclusion we may draw from the first table is that n-shot schemes are more susceptible to the Trojan Horse mechanism than one-shot schemes. To put it more
generally, the more divisible the scheme, the more effective the Trojan Horse mechanism. From the criminals’ point of view, this conclusion suggests a way of lowering the effectiveness of the Trojan Horse: they should try to design less divisible schemes. This, however, is often quite a difficult tactic to take. Conspirators may be able to postpone capturing the ill-gotten dividends to the end-game, but it seems almost impossible, or at least very expensive, to avoid producing incriminating evidence along the way.

A more practical method that criminals may utilize to deal with the divisibility problem is to move from finite to infinite schemes. Comparing n-shot finite schemes and infinite schemes, we find that the minimum fine required to deter infinite schemes is significantly higher. Criminals, therefore, in response to the introduction of the Trojan Horse mechanism, could be expected to devise their schemes with no finite horizon. Hence, although the Trojan Horse mechanism erects entry barriers to the crime market, at the same time, it might strengthen organized crime.6

D.3. Trojan Horses versus Bounty Hunters: Principal Distinctions

On first impression, it seems that the Trojan Horse enforcement mechanism is only a unique example of a bounty hunter program. Arguably, there is nothing in Becker & Stigler’s private enforcement model that denies co-conspirators the right to snitch and collect the fine. Similarly, some bounty hunter programs do not rule out the possibility of rewarding a squealing co-conspirator. But Becker & Stigler’s model as well as the prevailing bounty hunter programs do not provide adequate incentives for co-conspirators to come forward, as they fail

---

6As shown above, however, the Trojan Horse mechanism motivates criminals to reduce the number of accomplices in each scheme. In this respect, the Trojan Horse mechanism is effective against organized crime.
to take into account the *Trojan Horse*’s significant filing costs and, more importantly, his opportunity costs.\(^7\)

Unlike bounty hunters, a *Trojan Horse* plaintiff waives the benefits of the illegal scheme. Hence, for the *Trojan Horse* mechanism to kick in, the law must ensure *Trojan Horses* a much larger reward than the reward it provides to any other private enforcer. In fact, as shown in the two tables above, it is often the case that the *Trojan Horse* mechanism is effective only if we allow *Trojan Horses* to collect the entire fine the defendant pays. As shown above, ensuring a very large reward to the plaintiff-conspirator is particularly essential in infinite schemes.

In practice, however, even bounty hunter programs that do allow law-breakers to squeal and collect a portion of the fine do not provide any beneficial treatment to co-conspirators. On the contrary, relative to a law-abiding squealer, co-conspirators are typically awarded a lesser portion of fine.\(^8\) Under these circumstances, it seems unlikely that conspirators would snitch on their partners before the police is on their trail.

### D.4. Trojan Horses and State Witness Programs: Principal Distinctions

---

\(^7\) Another important distinction between the prevailing bounty hunter programs and the mechanism this paper endorses can be found in the role played by public enforcement agencies. Whereas in this paper, the enforcement agencies stay outside of the loop, the prevailing bounty hunter programs allow regulators discretion as to whether and to what extent to reward a snitch. Thus, although some laws empower regulators to reward co-conspirators for snitching, these regulators typically refuse to do so. For an excellent review on this subject, see M.J. Ferziger & D.G. Currell, *Snitching for Dollars: The Economics and Public Policy of Federal Civil Bounty Programs*, 1999 Univ. Ill. L. Rev. 111 (1999).

\(^8\) See Ferziger & Currell, *Id.*
State witness programs provide amnesty to conspirators who snitch on their partners. Hence, these programs, more than bounty hunter programs, resemble the *Trojan Horse* mechanism.

One important, institutional distinction between the *Trojan Horse* mechanism and amnesty programs is that amnesty programs are operated by a public enforcement agency who gets to decide whom, if any, to sue and to whom, if any to grant immunity. The *Trojan Horse* mechanism, on the other hand, is monitored by courts only and is not subject to the politics of prosecutorial discretion. Moreover, unlike prosecutors, whose decision-making process is typically confidential, court proceedings are open to public scrutiny. Thus, criminals are more likely to trust courts.

The second distinction between the two mechanisms is that whereas state witness programs only grant immunity from sanctions, the *Trojan Horse* mechanism offers a carrot as well. Thus, conspirators would approach a public enforcer and ask for immunity only after the public enforcer is on their tail; *Trojan Horses*, on the other hand, would snitch even with no public enforcers around. This distinction is of particular significance in infinite schemes. To see why, consider the case of an infinite scheme in which each conspirator anticipates a stream of revenues with a net present value of $G$. The conspirators also know that there is a probability $P$ they will be detected by the public authorities and forced to pay a fine with a net

---

9 The Antitrust Amnesty Program seems to come close to waiving prosecutorial discretion, as it ensures immunity to the first cartel member to come forward. See DOJ Corporate Leniency Policy (Issued August 10, 1993). However, even this program discriminates between conspirators on “moral grounds” and would not grant immunity to “the leader in, or originator of, the activity.” (Section 6 of the DOJ’s policy). Such a condition undermines the efficacy of the program because the minor conspirator is never threatened by the possibility that the main wrongdoer will sue first.
present value of $F$. For each conspirator, therefore, the net present value of the scheme is $G - PF$. As long as the net present value of the scheme is negative, public enforcement provides sufficient deterrence and there is no essential need for supplementing it with an amnesty program. The interesting case, therefore, is when the net present value of the scheme is positive, namely, when $F < G/P$. In such cases, it seems that an amnesty program cannot provide sufficient motivation for conspirators to snitch on one another. Even if filing costs are zero, as long as the net present value of the scheme is positive, it is worthwhile for the conspirators to continue engaging in the infinite scheme and it does not pay to squeal. Only a positive reward, like the reward this paper suggests providing to a Trojan Horse plaintiff, has the potential of tipping the scales and persuading participants in a profitable scheme to incriminate their partners. In other words, whereas amnesty programs are effective only against infinite schemes with negative net present value ($NPV$), the Trojan Horse mechanism deters profitable schemes as well.

D.5. Opportunism

One of the most troubling aspects of the Trojan Horse mechanism is that it might provoke opportunistic behavior. Consider, for example, the following hypothetical: The general belief among producers in a certain market is that an agreement among competitors to set quality standards does not violate the antitrust laws. Although such an agreement would clearly benefit these producers, free-riding predicaments entail that not one of them is willing to sustain the costs of initiating negotiations for such an agreement. Now suppose that one of these competitors suddenly learns about the illegality of such a standardization agreement. Assuming this informed party believes that the other producers are uninformed, the Trojan
enforcement mechanism might tempt that informed party to invite his naive competitors to enter into such an illegal scheme.

Many would find this effect troubling even if the opportunistic Trojan Horse were to file suit at the very earliest stage of the scheme, before any harm is done. Most troubling is the fact that the law puts together a mechanism that encourages otherwise law-abiding citizens to join or even initiate an illegal scheme. Arguably, they would do so for the sole purpose of filing a suit and collecting the fine from their “naive” accomplices.

One comforting answer to this objection is that the risk of such opportunism is probably not significant. The models I have used above missed this point because they assumed criminals conspire contemporaneously. But a seducer must always make the first illegal move and, thus, the seducer’s risk of being sued is greater than the second-mover’s risk. In our price-fixing example, the knowledgeable party can rarely be sure that all his competitors are, in fact, ignorant of the illegality of the scheme. Given the fact that the seducer must make the first illegal move, he faces the risk of being sued by one of his competitors, for the mere attempt to fix prices, before he finds his way to the courthouse.10

Another comforting answer to this concern is that opportunism might yield certain socially beneficial by-products. One side effect of the fact that the Trojan Horse mechanism

10This answer is not satisfactory in asymmetric cases like the illegal employment example discussed in the next part. With no money in his pockets, an immigrant can search for illegal employment with no risk of being sued. Note, however, that in the context of illegal employment, we can probably assume that the deep-pockets are informed, sophisticated agents and, therefore, are least likely to be entrapped.
may motivate (otherwise) law-abiding agents to entrap others is that it encourages agents to search for legal information and to publicize this information. Currently, the search for legal information is encouraged by the sanctions the law imposes on law-breakers. With the Trojan Horse mechanism in the background, potential law-breakers might reap profits from legal information. Some parties would search for such information for the purpose of entrapping other agents, while others will do so to protect themselves from the former. The market for legal information, therefore, would certainly benefit from this mechanism.

If, despite these two encouraging answers, we are still interested in restricting such opportunism, we can adopt a variant on the first-felon-to-file-collects-the-fine rule that would mitigate this problem. One radical solution would be to deny standing to anyone who knowingly conspires to violate the law. Under this solution, only innocent parties, who realize they are engaging in illegal conduct only after it has commenced, would be allowed to bring such a suit. The flaw in this solution, however, is that although it might operate effectively to deter opportunistic behavior, at the same time, it would undermine the efficacy of the Trojan Horse mechanism. For example, one possible scenario would be that the briber and bribee commit not to sue one another simply by declaring, in their bribery agreement, the illegality of their interaction.

Hence, to mitigate opportunism and, at the same time, to sustain the efficacy of this mechanism, we must ignore the plaintiff’s motivation and focus on the defendant’s state of mind. One logical refinement would be to subject Trojan Horses to the same limitation any police informer is subject to under the entrapment doctrine.11 Thus, the right to collect the fine

11Model Penal Code, Section 2.13.
would be contingent on the plaintiff’s ability to demonstrate that the defendant would have engaged in the illegal scheme even without the plaintiff’s involvement. Although this refinement does not eradicate opportunism, it does reduce the probability that the legal mechanism erected for the purpose of diminishing illegal schemes would, in fact, produce such schemes.

D.6. The Problem of False Claims

Another concern the Trojan Horse mechanism provokes is that felons would file false complaints. The prize the mechanism offers might induce felons to file false claims against innocent parties. Although I admit that the possibility of false claims is, indeed, troubling, I believe that the risk of a false Trojan Horse claim is no greater than the risk of a frivolous tort claim.

Similarly, if the Trojan Horse mechanism were to operate in tandem with public prosecution, the immunity it would provide to a co-conspirator plaintiff should not elicit more false complaints than an amnesty or a state witness program does. Actually, unlike state witnesses, a Trojan Horse who fails at trial is not immune from public prosecution, and the evidence he produces as a plaintiff is likely to be used against him when the public enforcement agencies go after him. Hence, assuming the risk of failure is higher when the claim is false, the risk of false Trojan Horse claims seems to be small, at least relative to the risk of false testimony by a state witness or a bounty hunter.
III. ASYMMETRIC GAMES AND THE PROBLEM OF FINANCIAL INADEUACY

Any enforcement mechanism that is based on economic sanctions alone seems to be ineffective when utilized against judgment-proof defendants. Consider again the case of illegal employment. Economic sanctions may deter employers, but the employees’ empty pockets clearly protect them from economic sanctions. It seems that the only way to discipline judgment-proof law-breakers, like illegal immigrants, is through jail sentencing and deportation. Experience indicates, however, that the opportunity costs for judgment-proof felons are often so high that they engage in the illegal conduct even in the face of a very high risk of jail sentencing and/or deportation. Moreover, subjecting only judgment-proof defendants to physical sanctions like a jail sentence, or deportation in the case of illegal immigrants, is a very costly move for the government, mainly because of the human rights sentiments it arouses.

This brings us back to the advantages of the *Trojan Horse* mechanism. The argument that economic measures do not deter illegal employees and other judgment-proof felons fails to recognize the fact that economic sanctions *can* work against judgment-proof law-breakers if they are presented as carrots rather than sticks. When one party to an illegal scheme is judgment-proof and the other has adequate financial means to meet the fine, the *Trojan Horse* strategy operates as a stick against the deep-pocket party and as a carrot for the judgment-proof party. In the illegal employment example, the *Trojan Horse* strategy provides adequate deterrence against employees, just as it does against employers, only against the former it uses carrots instead of sticks: it encourages illegal employees to sue and collect the fine from their employers, thereby ending the employment relationship. Similarly, allowing illegal
immigrants to collect the fine only on their way out of the country will encourage them to leave voluntarily.

The following model demonstrates few of these contentions.

**A. AN INFINITE ASYMMETRIC CONSPIRACY MODEL**

Consider an asymmetric game, like illegal employment, where conspirators have different levels of financial resources and they engage in the illegal conduct for the long–run, with no endgame on the horizon. For simplicity, I will assume that one party has no financial resources to pay the fine and the other is a deep pocket. I will further assume, as often is the case, that the empty-pocket criminal derives \$m from each incident of law-breaking and the deep-pocket derives \$g, where \( g > m \). Interest rate is still \( r \), so that \( m/r=M \) and \( g/r=G \). If the empty-pocket sues and proves his case, the deep-pocket pays a fine of \$F, a portion of which, \$R (\( R \leq F \)), is allotted to the empty-pocket and the rest goes to the government. The deep-pocket conspirator may also sue, but only for the purpose of protecting himself from a counter-suit by the empty-pocket conspirator. Filing costs for the deep-pocket and for the empty-pocket are \$D and \$E,\textsuperscript{12} respectively. The probability of winning a case remains \( Q \) for both conspirators. Under this setting, the infinite game looks as follows:

\textit{Table D: The Infinite Asymmetric Conspiracy Game}

<table>
<thead>
<tr>
<th>Deep-Pocket’s Payoff</th>
<th>Deep-Pocket Cooperates</th>
<th>Deep-Pocket Sues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty-Pocket’s Payoff</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{12}I assume that empty-pocket conspirators are still able to bear the filing costs, because as long as the net present value of filing a \textit{Trojan Horse} complaint is positive, empty-pocket conspirators should be able to find a financier for the suit.
The Empty-pocket would sue if:

\[(26) \quad QR-E > M\]

The deep-pocket would sue only if he knows that the empty-pocket is about to sue and only if:

\[(27) \quad QF > D\]

We are left to make sure that the parties would not want to engage even in one violation of the law. The empty-pocket would engage in one incident of law-breaking and sue if:

\[(28) \quad m + QR-E > 0 \quad \text{and} \quad QF > D\]

or if:

\[(29) \quad m + (QR-E)/2 > 0 \quad \text{and} \quad QF < D\]

Since we assume \(QR-E > M\) (condition 26), condition (28) and (29) always are satisfied. Hence, we may conclude that the empty-pocket will be interested in engaging in one incident of law-breaking for the purpose of filing a suit if:

\[(30) \quad QR > M + E\]

Turning to the deep-pocket, we know that he would engage in one violation only if:

\[(31) \quad g - QF > 0 \quad \text{and} \quad D > QF\]

or if:

\[(32) \quad g - (D + QF)/2 > 0 \quad \text{and} \quad D < QF\]

We may conclude, therefore, that the deep-pocket would not engage in the first law-breaking incident if:
Solving for the minimum fine that would meet conditions (30) and (33), we find that:

\[(34) \quad F^* = R^* = (M+E)/Q \text{ and } D \geq g\]

B. ANALYSIS

One lesson condition (34) teaches us is that the minimum fine that would deter asymmetric schemes is a function of the gains the empty-pocket conspirator anticipates to receive, rather than those anticipated by the deep-pocket. This result is in sharp contrast to the fines that public enforcement agencies should impose in such cases, which are a function of employer’s gains. Thus, if filing costs for empty-pocket conspirators are low enough, the Trojan Horse mechanism would be more effective than public agencies even if the probability of success for public enforcement is as high as that of Trojan Horses. For similar reasons, comparing condition (34) with the minimum fine for infinite symmetric schemes, we find that the minimum fine in asymmetric schemes is lower. We may conclude, therefore, that the Trojan Horse mechanism is more effective in asymmetric schemes.

Notice, however, that while setting the fine and the reward according to condition (34) above would ensure that employers are deterred from hiring illegal immigrants, it does not place employers and illegal immigrants in the conspirator dilemma. As long as the filing costs for employers are lower than their expected fine \((D>FQ)\), an employer who mistakenly hired illegal immigrants is better off continuing the employment relationship even if he knows the employee is likely to sue. Knowing this, an illegal immigrant that has been hired is not under
any pressure to file a suit as early as possible. We may mitigate this problem by several means.

First and most obviously, we may preserve the threat of the traditional enforcement mechanisms. Similarly, the law could employ Becker & Stigler’s mechanism and place the right to collect the fine in the hands of the first citizen to file a suit against the employer. Such potential competition from public or private enforcement agents would put illegal employees on notice that if other enforcers sue their employer before they do, they will lose their right to collect the fine from their employer. This threat would encourage illegal employees to chase the carrot and bring an end to their illegal employment at the very outset of the employment relationship.

Second, the state may promise the financially adequate conspirator a reward for putting his empty-pocket co-conspirators behind bars. More specifically, in our example, an employer would be rewarded for filing a suit for the imprisonment or deportation of his illegal employee. The employer’s right to collect the reward would be conditioned, of course, upon filing the suit before the illegal employee does so.

Lastly, there could be a more elegant solution, one that would force the Trojan Horse employee to gallop faster, but still preserves the spirit of the Trojan Horse mechanism. Conceptually, we could subject illegal employees to competition amongst one another by awarding the first illegal employee to come forward with the right to file a class action suit

---

13 The fact that the government rewards a criminal should not overly concern us. The government currently provides immunity and even promises financial reward to suspects who spill the beans and turn against their collaborators. The Trojan Horse plaintiff renders a more beneficial service to society, as he turns in criminals.
against his employer on behalf of all the illegal employees of the defendant. The larger the piece of the pie the law places in the hands of the class representative and the smaller the portion of the passive class members, the stronger the pressure on the employees to sue.

C. OPPORTUNISM

Deep-pockets have nothing to gain from entrapping judgment-proof agents in an illegal scheme. Judgment-proof felons, on the other hand, face no risk of being sued and, therefore, can only gain by seducing deep-pockets into a conspiracy. Thus, although the opportunism problem is one-sided, it may still be more prevalent in asymmetric schemes. However, one should bear in mind that typically, judgment-proof felons are also the less informed than deep-pockets. Employers, for example, are typically more familiar with their legal risks and rights than (potential) illegal immigrants, who may learn about their right to sue their employer only after they immigrate and start working. Moreover, an attempt on the part of a potential illegal immigrant to seduce employers involves significant costs – traveling costs, for example – and it is unlikely that someone would incur these costs only for the hope that he will find an uninformed employer to fall into the trap.

The problem of opportunism is of least concern in the most typical of the asymmetric schemes, where the deep-pocket party conspires with more than one judgment-proof agent. In such cases, it would be relatively easy for a judgment-proof conspirator to meet the requirements of the entrapment doctrine by demonstrating that the deep-pocket agent would have engaged in the illegal scheme even without his solicitation. For example, an illegal
immigrant who was recruited for the job in his home country would probably have no difficulty in proving that his employer would have hired an illegal immigrant even without solicitation.

IV. SUMMARY
This paper proposes a new enforcement mechanism against conspiracies, one that relies solely on the conspirators themselves. According to my proposal, the law should place conspirators in the conspirator dilemma by granting each one the right to sue the others. The first conspirator to file will collect the fine, or a portion of the fine, that the defendant pays. Using a simple model, I have analyzed the conditions under which the Trojan Horse mechanism is effective and compared its efficacy with that of alternative enforcement mechanism like public enforcement agencies, bounty hunters, and amnesty programs.

One important conclusion of this paper is that the Trojan Horse mechanism is particularly effective in asymmetric cases, where one party to the conspiracy is judgment-proof and the other a deep-pocket. In the context of such conspiracies, it is also easier to justify the Trojan Horse mechanism on moral grounds. Whereas in the bribery example, a moral argument can be brought against rewarding someone for entering into a bribery agreement, very few people would argue that there is something immoral about illegal immigration. Immigration law is not based on moral considerations, but on society’s preference of its own citizens over foreigners. In fact, even proponents of the criminalization of illegal immigration, prostitution,

14 This argument also is inaccurate: the reward is not given for participation in the crime, but, rather, for putting an end to the scheme.
or drug-use would concede that from a moral perspective, the violators of these laws are actually the victims. In such cases, therefore, any legal system that would put an end to these activities without using criminal sanctions will be generally considered morally superior and politically conceivable. The *Trojan Horse* mechanism provides such a solution.