Manipulating Reviews in Dark Net Markets to Reduce Crime

Panos Markopoulos  
Department of Business and Public Administration, University of Cyprus, markopoulos.panos@ucy.ac.cy

Dimitris Xefteris  
Department of Economics, University of Cyprus, xefteris.dimitrios@ucy.ac.cy

Chrysanthos Dellarocas  
Department of Management Information Systems, Boston University, dell@bu.edu

Online marketplaces account for a rapidly growing portion of the global trade in illicit products and services. Platforms like Agora and Silk Road have brought together three key technological innovations – anonymous Internet browsing through TOR, anonymous payments through electronic currencies, and the ubiquitous consumer reviews mechanism – and have built Dark Net Markets (DNMs) that have proven resilient to efforts by Law Enforcement Agencies (LEAs) to interfere with their actual operation. In this paper we ask whether LEAs may be able to interfere with the effective operation of DNMs, by manipulating buyer reviews in such a way so as to meaningfully reduce the volume or value of goods being traded. Our game-theoretic analysis shows that a LEA is not able to target the volume of goods traded. This happens because the buyers are able to fully account for LEA’s actions, and the LEA fails to have an impact. However, the LEA is able to successfully target the value of the goods traded. In this case the LEA is always able to reduce the informativeness of the reviews, as the optimal LEA strategy when targeting trade volume always takes into account product quality, which the buyers cannot fully account for, due to inherent quality uncertainty. The LEA is not only able to reduce the quality of the goods traded, but also their value and seller profitability, and moreover, to do so with a positive utility for itself.

Key words: online product reviews, review manipulation, online crime

1. Introduction

Only a few years ago, it would be hard to imagine that large volumes of highly illegal transactions would be arranged for in the open, right before the eyes of Law Enforcement Agencies (LEAs),
who are nevertheless powerless to prevent them from completing. Yet, this is exactly what online
marketplaces, such as *Agora* and *Silk Road*, have achieved. In this paper, we will use the term Dark
Net Markets (DNMs) to describe these marketplaces.

In the past, marketplaces of illicit goods were highly restrictive about participation. Restrictions,
such as secrecy, vetting, and the use of cash (which requires physical presence), where necessary
to keep LEAs at bay, but restrictions also potentially reduced transaction volume by effectively
screening out many buyers and sellers who would otherwise be willing to transact. Such restrictions
are now being pushed aside by electronic marketplaces that bring together three key technological
ingredients: anonymous browsing, electronic currencies, and consumer reviews. These marketplaces
maximize participation by keeping their doors open to anyone from around the world. At the same
time they appear to reduce the probability that participants will face legal consequences for their
actions.

The first widely known online bazaar of this kind, *Silk Road*, started operations in 2011 (Christin
2013). Competitors were quick to emulate it. In August 2014, the 5 biggest such markets, featured
more than 40,000 listings in the drugs category alone. Not only this represented a 50% increase in
the number of listings since the beginning of the year, but the markets appeared to be diversifying
in other product categories as well, with the largest player *Agora* also introducing powerful semi-
automatic firearms (Greenberg 2014a).

DNMs do not themselves sell anything. Instead they bring together buyers and sellers, and
usually charge a transaction fee. While individual details differ, the general process involves a
buyer signing-up by completing minimal information such as a username and password\(^1\). The buyer
then browses and chooses products in a manner that closely resembles mainstream marketplaces,
including distinct product categorization, access to product and seller reviews, and add-on services
(usually designed to provide increased anonymity and security for a given transaction). In case of

\(^1\) Some of DNMs are by invitation only, but such invitations are widely considered easy to obtain, and are often
reusable (Greenberg 2014a).
products that involve shipping, the buyer may submit his desired delivery address encrypted with the seller’s public cryptographic key to ensure that only the seller can see it. The buyer’s payment is not directly transferred to the seller but is channeled to an escrow account, usually managed by the DNM. When the buyer considers the transaction complete he/she finalizes the sale by asking the escrow service to release payment to the seller, and completes a (usually mandatory) seller review.

During the entire process, buyers and sellers use the Tor network (Dingledine et al. 2004; Fowler 2012), which is accessible with freely distributed software. Tor channels messages through a large number of network nodes called onion routers (Reed et al. 1998), and effectively hides the identity of its users from LEAs that may be observing network traffic. Payments are made using cryptocurrencies, such as Bitcoin (Kroll et al. 2013), providing an extra layer of anonymity.

Reached through anonymizing networks, and comprised of pseudonymous participants who communicate securely via public-key cryptography and who settle payments via crypto-currencies, DNMs have proven remarkably resilient to efforts of LEAs to interfere with their operations by the use of standard policing techniques. For instance, in late 2013, USA Federal Agents apprehended the alleged owner and operator of the original Silk Road marketplace, after a misconfiguration in the marketplace server leaked the server’s true location (Greenberg 2014b). However, no further arrests were made, as the information contained in the seized server could not help in the identification of other DNM participants. In fact, in a matter of a few weeks Silk Road 2.0 became operational (it appears that this was done with the help of some of the original Silk Road’s administrators), and the new DNM promptly exceeded its predecessor in the number of drug listings (Greenberg 2014a). Online sources that aggregate relevant info, such as the “SilkRoad” subreddit in the popular Reddit community, or the www.gwern.net website, suggest that the number of sellers or buyers of DNMs who have been apprehended so far, has been quite small, with arrests occurring often

---

2 This was later repeated with other DNMs. The DNM was either able to resurface shortly after, or was simply supplanted by another one.
after traditional police “sting” operations (Nark 2013), or after LEAs take advantage of technical errors or protocol vulnerabilities, which are then promptly patched by market operators.

Operating with their participants under anonymity, the unregulated black markets are held together by the review mechanism. Without the mechanism most buyers would understandably feel reluctant to trust sellers who they have no other way of contacting, and who are demonstrably willing to break rules and laws, if given the chance. DNM operators know this well: they often require that all buyers post a seller review after a transaction is complete, and make it costly for sellers to start again with a new identity by requiring that all sellers post a bond before allowing to sell.

In this paper we focus on the review mechanism. We ask whether LEAs can manipulate reviews to reduce the volume and/or the value of the goods transacted, and if so, how. By using a game-theoretic model, we find that it is indeed possible for LEAs to use manipulation of consumer reviews in order to attack DNM operations, but they should target the value of the goods transacted, not the volume.

Our results have important practical implications. LEAs’ efforts to attack participant anonymity (Ball et al. 2013) and traceless payments (Reid and Harrigan 2012), in order to discourage market operation have so far proven largely fruitless. This is not surprising, as Tor and crypto-currencies were designed exactly to thwart coordinated efforts to de-anonymize them. Any flaws discovered in these technologies, are quickly patched, so that LEAs may not be able to use similar attacks in the future. On the other hand, the review mechanism that the DNMs currently employ is designed to prevent transacting parties from cheating one another. At its heart, the review mechanism assumes that buyers will try to minimize the cost of a given purchase, and that sellers will try to maximize their profits. The mechanism was designed to prevent buyers and sellers from achieving these goals by cheating on other parties; it was not designed to prevent an outsider from reducing market activity, if that outsider was willing to spend resources for this purpose. This is exactly why attacking the review mechanism that the DNMs employ may succeed where other methods have failed.
2. A Model of a DNM

Our model is a modification of the model developed by Dellarocas (2006) to study strategic manipulation of reviews by sellers. We adopt Dellarocas’ notation in all model parameters that the two models share.

A monopolist firm sells a product that appeals to the buyers along two separate dimensions: one type (horizontal) dimension, and one quality (vertical) dimension. Buyer preferences for product type are uniformly distributed in an interval $[\delta_{\text{min}}, \delta_{\text{max}}]$ with unit density of buyers, per unit distance. We will assume that the interval is large enough so that we will not have to deal with boundary conditions. Buyers have linear transportation costs, and buyer $i$ gains the following utility from consuming a product that is $x$ units away from his ideal type:

$$u_i = V + q - r - x \cdot t$$  

(1)

where $V$ is the base buyer utility (the utility that buyers gain when product quality is zero), $q$ is the product quality, $p$ is price, and $t$ is the fit cost parameter (transportation cost) that we will normalize to $t = 1/2$. The risk-neutral buyers demand one unit of the good, subject to the constraint that their expected utility is positive.

Following standard theory, the Seller will locate his product at the center of the interval and draw demand $D = 4(V + q - p)$, accounting for buyers on either side of the product.

Variable costs are assumed to be zero, and thus profits are equal to sales revenues $w = D \cdot p$. If $q$ is common knowledge, then the Seller will maximize profit by setting $p = \frac{2q + V}{2}$. With known $q$ equilibrium product demand is then given by:

$$D = 2(q + V)$$  

(2)

for equilibrium demand, and

$$w = (q + V)^2$$  

(3)

for equilibrium profit.
Product quality $q$ is not known to the buyers. It is initially known only to the firm, and may become known to a LEA, who may purchase it and test it to ascertain its true quality. Buyers share a common prior about $q$. This prior is normally distributed with mean $m$ and precision (the reciprocal of the variance) $\tau$. In addition, buyers observe a normally distributed signal $x$ of the product’s true quality with mean $q$ and precision $\rho_x$. As in (Dellarocas 2006), we consider this signal to be the arithmetic mean of online ratings posted by buyers who have already purchased and received the product. The signal’s precision is then the sum of the precisions of individual ratings. While, on expectation, and in the absence of any efforts by LEAs, the review-related signal is unbiased, its realization may underestimate or overestimate the product’s true quality by an error term $\epsilon$ that depends on the precision $\rho_x$. Players do not know the realization of the signal $x$ a-priori.

Let $\theta = E(q \mid x)$ denote buyers’ posterior beliefs regarding the quality of the product, after they observe $x$. As in Dellarocas (2006), we will refer to $\theta$ as perceived quality. If the firm sets its price after buyers observe the signal $x$, Equations 2 and 3 yield:

$$D = 2(\theta + V)$$  \hspace{1cm} (4)

and:

$$w = (\theta + V)^2$$  \hspace{1cm} (5)

We assume that this market becomes the target of a LEA, which takes advantage of anonymity in order to post misleading reviews and manipulate the signal $x$. The LEA can shift the distribution of average ratings $q$ to $q + \eta$, where $\eta$ may be negative. To do this, the LEA is assumed to incur cost

$$c(\eta) = \lambda_{LEA} \cdot \eta^2$$  \hspace{1cm} (6)

This cost is assumed to consist of product purchases whose purpose is to give the LEA the ability to post reviews that will shift the quality signal by $\eta$. We must note three things about this cost. First, should the LEA decide to adopt this strategy, $c$ is added to the Seller revenue. Second, we
assume that the value of the goods that must be bought in order to influence the review signal, is
convex in the desired amount of manipulation. In other words, it gets progressively difficult for the
LEA to influence the review signal. More specifically, we assume a quadratic relationship, to make
our results comparable to the model by Dellarocas (2006), in which a quadratic cost is assumed
for manipulation by the Seller. Third, the parameter \( \lambda_{LEA} \) controls the unit cost of manipulation:
the lower \( \lambda_{LEA} \), the cheaper manipulation becomes, other things equal. We will take \( \lambda_{LEA} \) to be
exogenous, but we will discuss the market’s options to raise \( \lambda_{LEA} \), for example by demanding a
minimum purchase amount before allowing a buyer to review the product. Finally, we assume that,
should the LEA decide to incur the cost \( c \) and make a number of product purchases, it learns the
true product quality \( q \). For example, in the case of controlled substances, we assume that the LEA
performs the necessary laboratory analyses.

We will consider two different scenarios, that differ in the assumed LEA’s goals, and correspond-
ing utility functions. In the first case, the LEA attempts to reduce the value of the goods sold by
the Seller, or, since we assume zero marginal cost of production, to reduce the Seller’s profit. In
the second case, the LEA attempts to reduce the volume of the goods that reach buyers. We will
not argue about which strategy is better from a social-policy standpoint; we assume that each has
its merits. Instead we will compare them from an economic effectiveness standpoint.

2.1. LEA’s Options for Damaging Market Operations in the Continuous Model

In the first case the LEA targets the volume of trade. The LEA is willing to spend 1 dollar in order
to reduce the volume of the goods that reach the buyers by \( \beta \) units. LEA utility is given by:

\[
U_{volume} = \frac{E[D_b] - E[D_m]}{\beta} - \lambda_{LEA} \cdot \eta^2
\]  

(7)

where \( D_b \) is product demand in the absence of manipulation, \( D_m \) is product demand when the
LEA manipulates reviews, and \( \lambda_{LEA} \cdot \eta^2 \) is the value of the product purchases that the LEA must
perform to post misleading reviews and influence the quality signal by \( \eta \).

In the second case, the LEA is willing to spend 1 dollar in order to reduce Seller revenue by
\( \alpha > 1 \) dollars. Its utility is then given by:

\[
U_{value} = \frac{E[w_b] - E[w_m] - \lambda_{LEA} \cdot \eta^2}{\alpha} - \lambda_{LEA} \cdot \eta^2
\]  

(8)
where $w_b$ is base seller revenue, $w_m$ is seller revenue from non-LEA buyers when the LEA manipulates reviews. We now also account for the payment $\lambda_{LEA} \cdot \eta^2$ that the LEA makes to the buyer.

We assume that buyers cannot distinguish between real and LEA-generated reviews, but they are aware that a LEA is manipulating the market. Buyers rationally take this into account when forming their posterior beliefs about product quality (after observing the review signal). As in (Dellarocas 2006), we will calculate the Perfect Bayesian Equilibrium (PBE), where the LEA will maximize its expected payoffs given buyers beliefs about its manipulation strategy, and buyer beliefs will be consistent with the LEA’s strategy.

A general PBE would have the LEA choose $\eta$ according to an arbitrary function of quality: $\eta(q)$. Following Dellarocas (2006), we will focus on linear strategies of the form $\eta(q) = g + h \cdot q$, where $g, h$ are real numbers that on equilibrium match the buyers’ expectations about them. Buyers would then understand that, with LEA manipulation, the ratings they observe are the sum of the following indistinguishable components:

$$y = q + \eta(q) + \epsilon = q + g + h \cdot q + \epsilon$$

where $q$ is the true product quality, and $\epsilon$ is normally distributed with mean zero and precision $\rho_x$. Buyers would then be able to conclude that the true product quality is

$$q = \frac{y - g}{h + 1} - \frac{\epsilon}{h + 1}$$

so that the statistic $z = (y - g)/(h + 1)$ is a normally distributed unbiased estimator of $q$ with precision $\rho_z = \rho_x(h + 1)^2$. Buyers would then rationally update their prior for product quality $m$ to

$$\theta = \frac{\tau \cdot m + \rho_z(y - g)/(h + 1)}{\tau + \rho_z}$$

which we defined above as the product’s perceived quality.

2.2. LEA targets the value of trade

In the absence of manipulation, Seller revenue, from Equation 5 is

$$E[w_b] = E[(\theta + V)^2] = \frac{(\tau(m + V) + \rho_z(q + V))^2}{(\tau + \rho_z)^2} + \frac{\rho_x}{(\tau + \rho_z)^2}$$
Similarly, Seller revenue, when the LEA manipulates reviews is:

\[ E[w_m] = E[(\theta + V)^2] = \frac{(\tau(m + V) + \rho_z((q + \eta - g)/(h + 1) + V))^2}{(\tau + \rho_z)^2} + \frac{\rho_z}{(\tau + \rho_z)^2} \quad (13) \]

LEA’s objective is to choose \( g \) and \( h \) that maximize Equation 8. In the Appendix we show:

**Proposition 1.** There exists linear PBE where the LEA’s manipulation strategy is a linear function of the seller true quality \( \eta = g + h \cdot q \), where

\[ g = h \left( V + \frac{(m + V)\tau}{\rho_z} \right) \quad (14) \]

and \( h \) is a negative real solution of the polynomial equation:

\[ h = -\frac{\rho_z^2}{(1 + h)\lambda_{LEA}(1 + \alpha)(\tau + \rho_z)^2} \quad (15) \]

In Proposition 1, both \( g \) and \( h \) are always negative, so that \( \eta \) is always negative. This means that the LEA’s strategy would be to always give bad reviews, but more so when the product quality is high. In other words, the LEA would deflate the ratings of high quality sellers, more so than the ratings of low quality sellers.

Note the product \( \lambda_{LEA}(1 + \alpha) \) in the denominator of Equation 15. The unit cost of review manipulation \( \lambda_{LEA} \) is multiplied with the factor \( 1 + \alpha \) which depends on the relative value to the LEA of reducing Seller revenue. We can define the product \( \lambda_{LEA}(1 + \alpha) \) as LEA’s effective unit cost of manipulation.

It is interesting to compare LEA’s strategy with the strategy of a monopolist who manipulates reviews in order to maximize revenue, as is the case in Dellarocas (2006). To make things directly comparable let us assume that the LEA’s effective unit cost of manipulation in our model equals Seller’s unit cost of manipulation in (Dellarocas 2006). Let’s denote the \( g \) and \( h \) of Proposition 1 as \( g_{LEA} \) and \( h_{LEA} \), and the same parameters for the Seller as \( g_{Seller} \) and \( h_{Seller} \).

**Corollary 1.** The LEA’s strategy is exactly opposite to the strategy of a seller who manipulates reviews to maximize profit: \( g_{Seller} = -g_{LEA} \) and \( h_{Seller} = -h_{LEA} \), assuming that the seller has unit cost of manipulation that equals the LEA’s effective unit cost of manipulation.
**Proof:** We repeat the process used to prove Proposition 1 for a monopolist who faces unit manipulation costs $\lambda_{\text{Seller}} = \lambda_{\text{LEA}}(1 + \alpha)$, and who tries to maximize $E(w_m) - E(w_b)$. $\square$

In reality, we would not expect it to be the case that the unit cost of manipulation for the Seller would be as high as the effective unit cost of manipulation of the LEA. We would in fact expect that $\lambda_{\text{Seller}} < \lambda_{\text{LEA}} < \lambda_{\text{LEA}}(1 + \alpha)$. For example, if, as we assumed here, the most important cost component in review manipulation is the cost to purchase the product (earning the right to review), then, whereas the LEA would have to incur the full cost, the Seller would likely only have to incur the market transaction fee (usually 3-6% of product price) and otherwise fake the transactions. Even if we assume that the Seller and the LEA share the same unit manipulation costs, what Corollary 1 says, is that, because of the $1 + \alpha > 1$ factor, the Seller would be a more aggressive manipulator of reviews than the LEA, that is, the Seller would inflate his reviews more than the LEA would deflate them, for the same unit manipulation cost.

By manipulating buyer reviews, the LEA destroys the informativeness of the review signal and makes it less reliable. The reason why reviews are less informative when the LEA follows the strategy outlined in Proposition 1 is that $\rho_z = \rho_x(h + 1)^2 < \rho_x$, because $-1 < h < 0$ \(^3\); that is, the LEA reduces the precision of the review signal. By giving bad product reviews, but by doing more so when the product quality is high, the LEA brings together the means of the signal distributions $q + \eta$ that correspond to different qualities, thus making it more difficult for buyers to distinguish between high and low product quality. In other words, the LEA makes the distribution of the signal $y$ more crowded than the distribution of the review signal $x$, obtained in the absence of manipulation, making harder for buyers to tell apart high quality sellers from low quality sellers.

Making buyer reviews less informative is not always to the benefit of the LEA, nor does it hurt sellers regardless of quality. From Equations 10 and 11, the expected value for perceived quality without manipulation is

$$\theta = \frac{\tau \cdot m + \rho_x \cdot q}{\tau + \rho_x} \quad (16)$$

\(^3\) we need to explain why, see Dellarocas (2006) footnote 5
and with LEA manipulation perceived quality is

\[ \theta = \frac{\tau \cdot m + \rho_z \cdot q}{\tau + \rho_z} \] (17)

Perceived quality is the weighted average of the prior and signal means, weighted by the two precisions. When the LEA reduces the precision of the review signal by manipulating reviews, the buyers do not adjust their quality expectations as much as they would in the absence of LEA manipulation. This means that in the case of sellers whose quality is below the prior \( m \), LEA manipulation would cause buyers to fail to adjust their expectations downwards as much as they would with an honest review signal. These sellers would increase their revenue in the presence of review manipulation. On the other hand, sellers whose quality is above the buyer prior \( m \) would see their revenues reduce, as buyers fail to adjust their quality expectations upwards, as much as they would in the absence of LEA manipulation.

In the case of high quality sellers, the important question is whether the LEA is able to act profitably and come out with a positive utility, that is, whether the LEA is able to reduce the revenue of high quality sellers enough to justify the cost of its efforts. Proposition 2 says that this is indeed so, provided that the precision of honest ratings \( \rho_x \) is low enough. This is because the LEA’s impact on seller revenue decreases as \( \rho_x \) increases. Mathematically, as \( \rho_x \rightarrow \infty \), \( E(w_b) \) and \( E(w_m) \) both converge to \( q^2 \) and are not affected by the amount of review manipulation \( \eta \). Thus, we can easily show (proof omitted) that, if the precision \( \rho_x \) is high enough, the LEA cannot act profitably, even for sellers whose quality greatly exceeds the prior \( m \). However, when the precision of honest ratings is low enough, we can show:

**Proposition 2.** If the precision of honest ratings is low enough (\( \rho_x \) is below a threshold \( R_x \)) and the LEA values the reduction of seller revenue enough (\( \alpha \) is below a threshold \( 1 < \alpha < A \)), then, independently of the unit manipulation cost \( \lambda_{LEA} \), the LEA is able to profitably\(^4\) manipulate reviews of sellers of sufficiently high product quality

\(^4\) with positive utility
Arguably, the most important question that emerges is whether a LEA can expect to profitably manipulate buyer reviews unconditionally, that is, before learning the true product quality $q$. The answer is yes, if the manipulation costs are low enough and the prior $m$ does not systematically overestimate true product quality. The intuition is that, because seller profit is quadratic in quality, the seller revenue reduction in the case of a seller whose quality is above the $m$ prior by a given amount $\delta$, is greater than the revenue increase of a seller whose quality is below the $m$ prior by the same $\delta$. The proof of the following Proposition, is available from the authors, upon request.

**Proposition 3.** If the precision of honest ratings is low enough ($\rho_x$ is below a threshold $R_x$), the effective unit cost of manipulation $\lambda_{LEA}(1 + \alpha)$ is low enough, and the buyer prior about product quality $m$ does not systematically overestimate product quality ($m \leq q$), the LEA profitably manipulates reviews on expectation.

### 2.3. LEA targets the volume of trade

In the absence of manipulation, buyer demand, from Equation 4 is

$$E[D_b] = E[2(\theta + V)] = \frac{(\tau (m + V) + \rho_x (q + V))^2}{(\tau + \rho_x)^2} + \frac{\rho_x}{(\tau + \rho_x)^2} \quad (18)$$

Similarly, buyer demand, when the LEA manipulates reviews is:

$$E[D_m] = E[2(\theta + V)] = \frac{(\tau (m + V) + \rho_z ((q + \eta - g)/(h + 1) + V))^2}{(\tau + \rho_z)^2} + \frac{\rho_z}{(\tau + \rho_z)^2} \quad (19)$$

LEA’s objective is to choose $g$ and $h$ that maximize Equation 8. Following the process we used in the proof of Proposition 1, we show:

**Proposition 4.** There exists linear PBE where the LEA’s manipulation strategy is a constant function

$$\eta = -\frac{\rho_x}{\beta \cdot \lambda_{LEA} (\tau + \rho_x)} \quad (20)$$

independently of the seller true quality

$^5$with positive utility
Proof Taking the first order condition $\partial U_2 / \partial \eta = 0$ of Equation 7, and solving for $\eta$, we obtain $\eta = -\frac{(1+h)\rho_x}{\beta \lambda_{LEA}(\tau+(1+h)^2\rho_x)}$, which does not depend on quality $q$. For the solution to be consistent with buyers’ expectation of a linear $\eta = g + h \cdot q$, it must be $h = 0$. Thus $\eta = g = -\frac{(1+h)\rho_x}{\beta \lambda_{LEA}(\tau+(1+h)^2\rho_x)}$. □

Proposition 4 shows that when the LEA tries to reduce the volume of goods that reach buyers, it ends up deflating all reviews by a constant amount, irrespective of quality. This however does not reduce the informativeness of the review mechanism, as rational buyers can simply add the constant $\frac{(1+h)\rho_x}{\beta \lambda_{LEA}(\tau+(1+h)^2\rho_x)}$ to the ratings they observe to completely erase LEA’s impact to ratings. Indeed, setting $h = 0$ in Equations 9 and 11, we find $\theta = \frac{\tau \cdot m + \rho_x \cdot q}{\tau + \rho_x}$ which is simply Equation 16: the buyers’ perceived quality in the absence of LEA manipulation.

In this case the LEA would rather not spend any resources manipulating reviews, since it cannot reduce the volume of goods traded not even by a single unit, while it spends resources $\frac{\rho_x^2}{\beta^2 \lambda_{LEA}(\tau+\rho_x)^2}$ in the process. However, because the buyers expect that the LEA deflates product reviews, should the LEA decide to stay in the sidelines, buyers will erroneously include in their perceived quality $\theta$ the expected LEA review deflation, and product demand will increase, reducing the LEA utility function, given by Equation 7. The LEA will have to manipulate reviews just to maximize the (still negative) value of its utility function.

2.4. Targeting value versus targeting volume of trade

We saw that a LEA is not able to target the volume of trade and reduce it by manipulating reviews, as buyers are able to fully account for its actions. The strategy of targeting the volume of trade does not succeed in its goal, because it does not manage to reduce the informativeness of the review mechanism. However, the LEA is able to reduce review informativeness when targeting the value of goods traded. By manipulating reviews according to Proposition 1 the LEA allows the buyers to only partially account for the amount by which LEA deflates the reviews, because this amount depends on product quality, which is not perfectly known to buyers. Because the LEA deflates the reviews of higher quality sellers more so than lower quality sellers, when a buyer sees a review signal, he cannot know whether this is a high quality seller whose review has been greatly deflated, or a low quality seller whose review the LEA deflated only slightly.
An interesting question is whether the LEA reduces traded volume when it targets traded value, because we saw that in the latter case it manages to reduce the informativeness of the review signal. The answer is no. By targeting trade value, the LEA is able to transfer (on expectation) sales from high quality sellers to low quality sellers, reducing (average) seller profitability, but keeping traded volume constant. Any linear manipulation strategy that manages to increase review signal noise, including the equilibrium manipulation strategy employed when the LEA is trying to minimize seller profit, is unable to reduce the volume of traded goods. More generally, any such strategy is unable to reduce any metric that is linear in product quality (as it happens to be the case with traded volume in our model). On the other hand, any manipulation strategy that reduces the informativeness of the review signal may be able to reduce a metric that is convex in product quality (as it happens to be the case with the value of traded goods), if the manipulation cost is low enough\(^6\).

It thus appears that for a LEA that wants to cause damage to a DNM it is a more attractive to target the value of the goods, rather than the volume of the goods traded. When the LEA targets the volume of goods traded it fails to cause any other damage in the market: it fails to reduce seller profit, and it fails to reduce the average quality of the products traded in the market. On the other hand, when the LEA targets the value of the goods traded, it can achieve its goal and reduce seller profit. If the effective manipulation cost is low enough, it can even do so profitably, that is, it can reduce seller profit, while gaining positive utility from its actions. In addition, by targeting the value of traded goods it always manages to also reduce the quality of the products traded in the market, that may be an attractive side-goal for many LEAs.

3. Concluding Remarks

As trading of illicit goods and services moves to Dark Net Markets (DNMs) that protect the anonymity of participants and make their transactions untraceable, the ability of Law Enforcement

\(^6\)If the LEA targets a metric that is concave in product quality then it should try to increase, not decrease, the informativeness of the review signal.
Agencies (LEAs) to interfere with their operations by using traditional approaches seems severely limited. In this article we investigated the possibility that a LEA can instead target what may be the weak link that allows these markets to operate efficiently: the ubiquitous review mechanism that DNMs employ to build trust among market participants.

By employing a game-theoretic model, we studied the implications of a LEA manipulating buyer reviews in the DNM, in order to fulfill strategic goals, related to crime reduction. First we found that a LEA may be able to profitably (with positive utility) reduce seller revenue in a DNM, by following a strategy that deflates buyer reviews, by an amount that increases with product quality. This was shown to be possible when the effective unit cost of manipulation is low enough, for example if the DNM allows the LEA (disguised as a buyer) to review products even after making low-value purchases.

Second, we found that the LEA cannot target the reduction of traded volume, that is, it cannot target the reduction of the number of units of the product sold to buyers. The seller ends up deflating all product reviews by the same amount, a strategy that is easy for buyers to counter, by adding the exact amount back to the observed review signal. When the LEA targets the reduction of the volume of products traded, the buyers always force it to participate in a game that wastes its resources, without achieving any reduction in the volume of traded goods.

The most important implication of this work extends far beyond our specific findings. We have showed that, theoretically, the review mechanism that allows DNMs to operate efficiently may well be susceptible to manipulation by a LEA. Indeed, we showed that the review mechanisms currently employed (designed to protect market participants from each other) do not necessarily offer optimal protection from manipulation by LEAs whose objectives maybe markedly different than market participant objectives.

This opens up many possibilities to study other manipulation strategies as well. In our game-theoretic analysis we restricted to the study of linear Perfect Bayesian Equilibria. However, LEA strategies that aim to reduce the effectiveness of the review mechanism by randomizing review
ratings, or non-linear strategies that try to make all products appear indistinguishable from each other, may in fact prove even more efficient ways for LEAs to achieve their goals.

Appendix A: Proofs and derivations

**Proof of Proposition 1**

We first take the first-order condition \( \partial U_1 / \partial \eta = 0 \) of Equation 8, using Equations 12 and 13. Solving for \( \eta \), we verify that the Seller’s optimal strategy is a linear function \( \eta = g + h \cdot q \), with \( g = \frac{\rho_z (r + V) - \tau}{\lambda \alpha (\tau + \rho_z)^2 (h+1)^2 + \rho_z^2} \) and \( h = \frac{-\rho_z^2}{\lambda \alpha (\tau + \rho_z)^2 (h+1)^2 + \rho_z^2} \). In equilibrium, buyer beliefs are consistent with the optimal \( g, h \) chosen by the LEA. Solving the above equations for \( g, h \) we obtain Equations 14 and 15. The second order condition \( \partial^2 U_1 / \partial \eta^2 < 0 \) yields \( \frac{2\rho_z^2}{(\tau + \rho_z)^2 (h+1)^2} h < 0 \), so that only negative solutions for \( h \) are admissible. □

**Proof of Proposition 2**

As \( \alpha \) approaches 1, then, from Equation 8, \( U_1 \) approaches \( E(w_b) - E(w_m) - 2\lambda \eta^2 \). Because of Corollary 1, the difference in revenue is exactly the difference in revenue in the case where a seller is manipulating reviews to maximize profit. Further setting \( 2\lambda \eta = \lambda \cdot \eta^2 \), we only need to prove that, for low enough \( \rho_z \), the difference in revenue achieved by a seller who manipulates reviews to maximize profit, is always high enough to offset the \( \lambda \cdot \eta^2 \) factor, for sufficiently high product quality. This reformulation of our initial Proposition is proven in Dellarocas (2006). □

Appendix B: Key symbols used in the notation

<table>
<thead>
<tr>
<th>Key Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decision Variables</strong></td>
<td></td>
</tr>
<tr>
<td>( r )</td>
<td>Product price</td>
</tr>
<tr>
<td>( \eta(q) )</td>
<td>amount by which the LEA manipulates the reviews of a product of quality ( q )</td>
</tr>
<tr>
<td>( g, h )</td>
<td>the parameters of the linear function ( \eta(q) = g + h \cdot q )</td>
</tr>
<tr>
<td>Key Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>$V$</td>
<td>Base buyer utility (the utility that buyers gain when product quality is zero)</td>
</tr>
<tr>
<td>$m$</td>
<td>Mean of prior buyer beliefs about quality</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Precision or prior beliefs</td>
</tr>
<tr>
<td>$x$</td>
<td>Baseline review signal (average of honest buyer ratings)</td>
</tr>
<tr>
<td>$\rho_x$</td>
<td>Precision of baseline review signal (precision of $x$)</td>
</tr>
<tr>
<td>$y$</td>
<td>Observable review signal, including LEA’s manipulation</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>Precision of adjusted review signal, after LEA manipulation (precision of the unbiased estimator of $q$: $(y - g)/(h + 1)$)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Perceived quality (mean of buyer posterior beliefs about quality, after LEA manipulation)</td>
</tr>
<tr>
<td>$\lambda_{LEA}$</td>
<td>unit cost of manipulation (total cost is $c(\eta) = \lambda_{LEA} \cdot \eta^2$)</td>
</tr>
<tr>
<td>$q, D, w, U, \alpha, \beta$</td>
<td>Same as in the Binary Model</td>
</tr>
</tbody>
</table>

References


