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* and their subsequent imitators, 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 use a game theoretic model to study whether LEAs may be able to interfere with the effective operation of DNMs, by manipulating buyer reviews. Our main result is that a sufficiently well funded LEA can create "Market for Lemons" dynamics in the DNM, causing it to fail. However, a miscalculation on the part of the LEA can be expensive: if it attempts review manipulation with an insufficient budget, its efforts and resources may be completely wasted: the LEA may expend its entire budget without influencing the volume of goods sold in the DNM. A LEA can alternatively pursue secondary goals, such as the reduction of seller profitability, due to its detrimental impact on the review signal. However, even then the LEA must be careful: the funds that it injects into the market in order to manipulate the reviews, may cause a net increase in seller profitability. In a simple model with two quality levels, we are able to derive the optimal strategy for a LEA to achieve market failure.

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 many online marketplaces, from *Silk Road* to *Agora* and their current reincarnations, 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, such as powerful semi-automatic firearms (Greenberg 2014a). As of October 2015, there were about 30 DNMs in operation, with a daily revenue fast approaching $1 Million (www.gwern.com; Soska and Christin 2015).

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). In case of 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 *crypto-currencies*, 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

[1] Some of DNMs are by invitation only, but such invitations are widely considered easy to obtain, and are often reusable (Greenberg 2014a).

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)[2]. 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 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 or value of the goods transacted, and if so, how.* We introduce a simple binary model with two levels of product quality, and a simple review mechanism that a LEA can manipulate by incurring and appropriate review manipulation cost. We show that LEA manipulation makes it harder for the buyers to tell apart product qualities. We find that, given sufficient resources, the LEA can decrease the informativeness of the review mechanism to the point where the high quality seller will start behaving like a low quality seller, leading to a pooling equilibrium of low quality products, which in turn causes the market to collapse.

On the other hand, if the LEA does not have sufficient resources to bring the DNM to such a "Market for Lemons" (Akerlof 1970) tipping point, it expends its entire resources without affecting the volume of goods traded in the market: we show that as long as the high quality product is being produced, the market is resilient, with (less informed) buyers shifting their purchases to lower quality products at a lower price. The result is that the LEA expends resources without affecting the volume of goods traded in the market. It is nevertheless *stuck* in manipulating reviews, because

[2] This was later repeated with other DNMs. The DNM was either able to resurface shortly after, or was simply supplanted by another one.

the buyers expect it to do so. Should it stop, buyers would significantly increase their expectations for product quality which would lead to a further increase in the volume of products transacted.

In the case where the LEA cannot reduce the volume of goods traded, it may still wish to pursue secondary goals, such as the reduction of seller profitability. We find that, indeed, a LEA can reduce expected seller profitability, but only for certain ranges of the market parameters. Also accounting for the funds that the LEA injects into the DNM, the LEA can even cause a net increase in seller profitability.

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.

The mechanism of manipulation that we explore in this paper assumes that the LEAs post product reviews strategically. In publicly available FBI search warrant requests documented in *www.gwern.net* it is indeed revealed that LEAs do make purchases from DNMs, especially of controlled substances, for testing and other purposes. The Economist reported that the FBI made over 100 purchases on Silk Road before closing it down (Economist 2014). Along with these purchases, LEAs gain the ability to post buyer reviews in the DNM, but there is no evidence that they have taken advantage of this fact in order to inflict any damage to any DNM. In this paper we explore the possibility of LEAs making purchases in DNMs with the explicit goal of gaining the ability to post reviews strategically, in a manner that reduces DNM activity.

This paper is organized as follows: In Section 2 we introduce our simple model of a DNM and explore equilibria for a LEA who pursues its goals given a certain budget. After we study the market dynamics in detail, in Section 3 we introduce a utility-driven LEA that can cause market failure in an optimal (from an efficiency standpoint) manner. We proceed to make policy recommendations in Section 4, and conclude in Section 5.

2. A Model of a DNM

2.1. Market Operation and the Role of Reviews

A monopolist seller can be either one of two types. The seller can be of type "*High Quality*" S_H with probability f, or of type "*Low Quality*" S_L with probability $1-f$. Both types sell 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 $[\delta_{min}, \delta_{max}]$ 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:

$$u^i = q - r - x \cdot t \tag{1}$$

where q is the product quality, r 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.

Quality can take one of two different values: $q = 0$ or $q = 1$. Both seller types can produce at the low quality level $q = 0$, and the cost to do so is zero. In addition, the high quality seller S_H can choose to produce the high quality level $q = 1$, by incurring cost w.

Following standard theory, the monopolist will locate his product at the center of the interval. If quality were known, then the seller would draw demand $D = 4(q - r)$, accounting for buyers on either side of the product. Profits would be $D \cdot r - c$, where $c \in \{0, w\}$ is the quality production cost which is zero for the low quality product, and w for the high quality product. The seller would maximize profit by setting price

$$r = \frac{q}{2} \tag{2}$$

Equilibrium product demand would be:

$$D = 2q \tag{3}$$

and seller revenue:

$$R = q^2 \tag{4}$$

Note, that if quality is known to be zero, then product price, demand, and revenue are also zero. However, the buyers cannot observe q directly, prior to purchase. Quality q is initially known only to the seller, and as we will explain below, may become known to a Law Enforcement Agency (LEA) if it is willing to incur an appropriate cost. Instead, buyers observe a signal s, such that when $q = 1$ $(q = 0)$ the signal is $s = 1$ $(s = 0)$ with probability $p \in (\frac{1}{2}, 1)$ and the signal is $s = 0$ $(s = 1)$ with the complementary probability (See also the left side of Figure 1). Players do not know

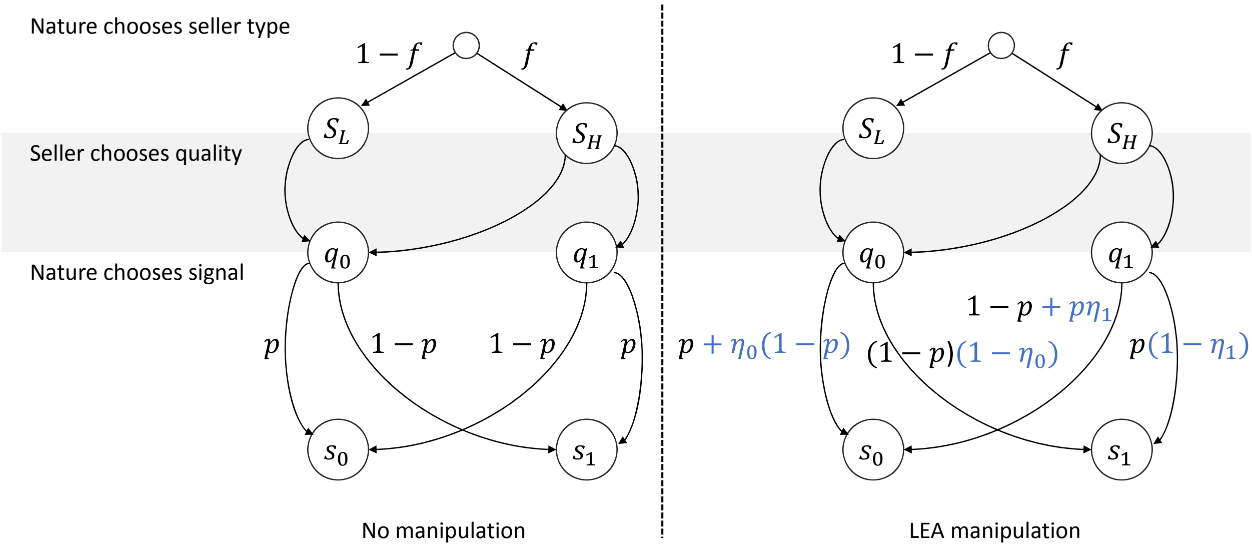

Figure 1 A Simple Model of a DNM. We denote $q = 0$ by q_0, $q = 1$ by q_1, and similarly for $s = 0$ and $s = 1$

the realization of the signal s a-priori. The signal is related to online ratings posted by buyers who have already purchased and received the product, and it can either be a "good-reviews" signal $s = 1$ or a "bad-reviews" signal $s = 0$. The probability p controls the accuracy, or informativeness, of the review mechanism. When $p = 1$ reviewers are always able to convey high or low quality information correctly to buyers, and when $p = 1/2$ the review signal is random and conveys no information about underlying product quality. Researchers have identified multiple reasons as to why reviewers may not all convey perfect quality information to buyers; many such reasons are discussed by He and Bond (2015).

Let $\theta = E(q \mid s)$ denote buyers' posterior beliefs regarding the quality of the product, after they observe s. If the firm sets its price after buyers observe the signal s, Equations 3 and 4 yield:

$$D = 2\theta \tag{5}$$

for product demand, and

$$R = \theta^2 \tag{6}$$

for seller profit.

The timing of the events, as described above, is also summarized in Figure 1. Specifically, we will restrict our analysis to pure strategy equilibria in all subgames. As in (Dellarocas 2006), we will derive the Perfect Bayesian Equilibrium (PBE), where the LEA will pursue its goals given buyers beliefs about its manipulation strategy, and buyer beliefs will be consistent with the LEA's strategy.

If the high quality seller is known to always produce the low quality good, then θ would be zero, and from Equation 5, $D = 0$. This is the case where the market fails to operate and support any transactions. However, if the high quality seller is known to always produce the high quality good, then θ is given by:

$$\theta_{b1} \equiv E[q \mid s = 1] = \frac{f \cdot p}{1 - p + f(2p - 1)} \tag{7}$$

and

$$\theta_{b0} \equiv E[q \mid s = 0] = \frac{f(1-p)}{p - f(2p - 1)} \tag{8}$$

Note that we use the subscript b in θ_b, to denote the *base case* or *baseline* where reviews are not manipulated. Also note that in the symmetric case $f = 1/2$ we have $\theta_{b1} = p$ and $\theta_{b0} = 1 - p$

Expected product quality, given q is:

$$E[\theta_b \mid q = 1] = p \cdot \theta_{b1} + (1 - p) \cdot \theta_{b0} \tag{9}$$

and

$$E[\theta_b \mid q = 0] = (1 - p) \cdot \theta_{b1} + p \cdot \theta_{b0} \tag{10}$$

We can now assess the role of the reviews in keeping the market operating. As we discussed above, market operation depends on S_H producing the high quality product. If S_H is known to produce the high quality product, then ex ante product quality is $f \cdot E[\theta_b \mid q = 0] + (1 - f) \cdot E[\theta_b \mid q = 1] = f$, which is independent of p and w. However, S_H expected profit is $E[\theta_b^2 \mid q = 1] - w = p \cdot [\theta_{b1}]^2 + (1 - p) \cdot [\theta_{b0}]^2 - w$ which depends on both p and w.

The consequence of the dependence of S_H profit on p and w is that there is a minimum required p_{min} and a maximum allowed w_{max} for the market to be able to operate. Beyond p_{min} and/or w_{max}, the high quality seller S_H would rather produce the low quality product, leading to market failure. S_H would only produce the high quality good if his expected profit exceeds the expected profit from producing the low quality good: $p \cdot [\theta_{b1}]^2 + (1 - p) \cdot [\theta_{b0}]^2 - w > (1 - p) \cdot [\theta_{b1}]^2 + p \cdot [\theta_{b0}]^2$, or

$$w_{max} = \left(\theta_{b1}^2 - \theta_{b0}^2\right)(2p - 1) \tag{11}$$

p_{min} is calculated by solving Equation 11 for p. Both w_{min} and p_{max} are depicted in Figure 2. In the special case where the prior f is 1/2 (S_H and S_L are equally likely), w_{max} reduces to $w_{max} = (2p - 1)^2$, and $p_{min} = \frac{1+\sqrt{w}}{2}$

As can be seen on the left side of Figure 2, w_{max} increases with p. The higher p is, the more the buyers trust the review signal, which makes it less likely that S_H will defect and produce the low

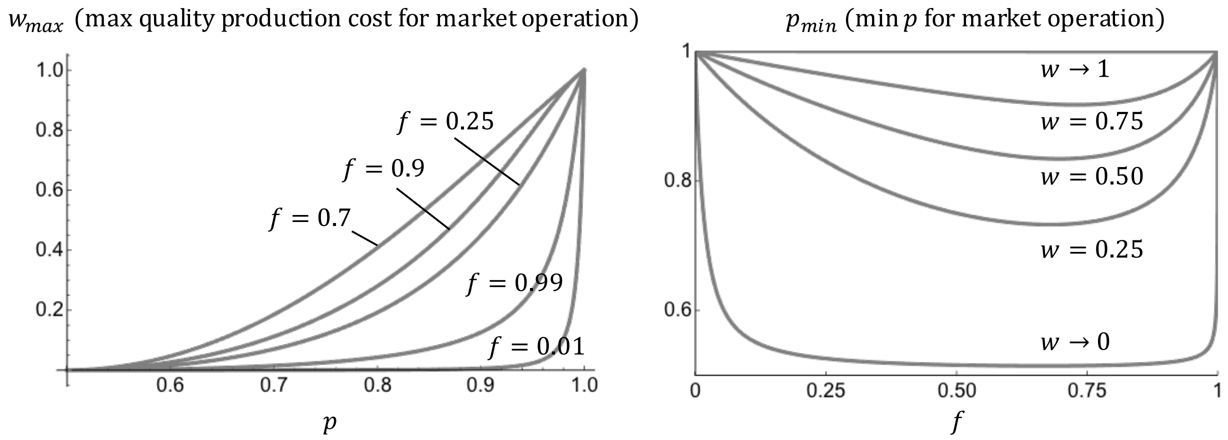

Figure 2 $\mathbf{w_{max}}$ **(left side) and** $\mathbf{p_{min}}$ **(right side), required for market operation.**

quality product, as buyers are then likely to find out. Thus, with higher p, a larger w is needed to justify defection.

We should also expect w_{max} to be lower for either very high, or very low f, as indeed the left hand side of Figure 2 shows. When f if close to zero or one, buyers tend to trust the strong prior more than the review signal. This makes it easier for S_H to defect, as the low quality signal that will likely result from the defection will have less impact to buyers. Defection by S_H will be more difficult (that is, it will require a larger w) for intermediate values of f as buyers will pay more attention to the review signal in that case.

The situation is similar for p_{min}, as the right side of Figure 2 shows. p_{min} increases with w, as higher quality production cost w makes defection more tempting, which necessitates a better review mechanism. However, p_{min} is not monotonic in f. As with w_{max}, when f if close to zero or one, buyers tend to trust the strong prior more than the review signal, and a strong review signal (high p) is needed to prevent S_H to defecting to low quality. When uncertainty about f is high (f is closer to 1/2) the review signal becomes important for buyers. Then, S_H is more careful to avoid the low-quality signal ($s = 0$), likely to be caused by defection.

2.2. Review Manipulation: Shutting Down the Market

We assume that the market becomes the target of a LEA, which takes advantage of anonymity in order to post misleading reviews and manipulate the signal s. The LEA can change the probability p that a high quality product results in a good-review by an amount $\eta = \eta_1$, and the probability $1-p$ that a low quality product can result in a good-review by an amount $\eta = \eta_0$. We specifically consider that when $q = 1$ the signal is $s = 1$ with probability $p(1-\eta_1)$ and $s = 0$ with the complementary probability, where η_1 is the amount by which the LEA is manipulating the reviews when $q = 1$. Similarly, when $q = 0$ the signal is $s = 1$ with probability $(1-p)(1-\eta_0)$ and $s = 0$ with the complementary probability, where η_0 is the amount by which the LEA is manipulating the reviews when $q = 0$. We depict this on the right side of Figure 1.

To do this, the LEA is assumed to incur cost $c(\eta) = L^2\eta^2$, where L^2 is some constant that reflects the difficulty of review manipulation, for a given level of η. 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 η. We must clarify 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, 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. This is why the LEA is able to choose specific η, either η_0 or η_1, based on the value of q. Finally, 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. The functional form of the manipulation cost is also appropriate when the DNM operator is actively seeking to undermine LEAs manipulation efforts.[3]

We do not restrict a priori η_0 and η_1 to be either positive or negative. The LEA is allowed to influence reviews both upwards and downwards. Conceivably, we could imagine a LEA posting positive reviews for low quality sellers in order to destroy the informativeness of the review mechanism. However as we see below, this is not an equilibrium strategy for the LEA who instead only gives bad reviews. The LEA, depending on its goals, may target high quality sellers for more aggressive manipulation.

Let $\hat{\eta}_0$ and $\hat{\eta}_1$ be the buyers' beliefs about η_0 and η_1 respectively. In other words, $\hat{\eta}_0$ and $\hat{\eta}_1$ are the amounts by which the buyers believe that the LEA will manipulate the review signal when $q = 0$ and $q = 1$ respectively

When the LEA is known to manipulate reviews, θ is given by:

[3] One way to interpret the cost $L^2\eta^2$ is as follows. Facing a "naive" DNM that does not actively seek to counter the LEA's actions, the number M of product purchases that would be required to reduce the average review of a seller with N "honest" reviews from q' to $q' - \eta$ is given by $\frac{q' \cdot N}{N+M} = q' - \eta \Leftrightarrow M = N\frac{\eta}{q'-\eta} \approx N\left(\frac{\eta}{q'} + \frac{\eta^2}{q'^2} + \frac{\eta^3}{q'^3} + \frac{\eta^4}{q'^4} + \dots\right)$, so that, as long as η is sufficiently smaller than q', the first two terms of the series are a reasonable approximation for the manipulation cost. Thus, given N and q, it is reasonable to assume that the cost grows with η^2. However, we would expect a DNM to actively defend against fake reviewers. In that case the LEA would only be able to safely manipulate reviews by posting a fake review, once every L honest ones. Then the number of product purchases that would be required to reduce the average review from q' to $q' - \eta$ would be given by $\frac{q \cdot N + M \cdot q' \cdot (1-1/L)}{N+M} = q' - \eta \Leftrightarrow M = N\frac{\eta \cdot L}{q' - \eta \cdot L} \approx N\left(\frac{L\eta}{q'} + \frac{L^2\eta^2}{q'^2} + \frac{L^3\eta^3}{q'^3} + \frac{L^4\eta^4}{q'^4} + \dots\right)$ Thus, given N and q, it is reasonable to assume that the cost grows with $L^2\eta^2$, as long as $L^2\eta^2$ is sufficiently smaller than q'.

$$\theta_{m1}(\hat{\eta}_0,\hat{\eta}_1) \equiv E[q \mid s=1] = \frac{p(1-\hat{\eta}_1)f}{(1-p)(1-\hat{\eta}_0)(1-f)+p(1-\hat{\eta}_1)f} \tag{12}$$

and

$$\theta_{m0}(\hat{\eta}_0,\hat{\eta}_1) \equiv E[q \mid s=0] = \frac{((1-p)(1-\hat{\eta}_1)+\hat{\eta}_1)\,f}{(p(1-\hat{\eta}_0)+\hat{\eta}_0)\,(1-f)+((1-p)(1-\hat{\eta}_1)+\hat{\eta}_1)f} \tag{13}$$

where we use the subscript m in θ_m, to denote that the reviews are being manipulated. Note that $\theta_{b0}=\theta_{m0}(0,0)$ and $\theta_{b1}=\theta_{m1}(0,0)$. We will also be using the shorthand θ_{m0} and θ_{m1} for $\theta_{m0}(\hat{\eta}_0,\hat{\eta}_1)$ and $\theta_{m1}(\hat{\eta}_0,\hat{\eta}_1)$ respectively.

Consequently, when the high quality seller is known to always produce the high quality good, expected product quality, given q is given by:

$$E[\theta_m \mid q=1] = p(1-\eta_1)\cdot\theta_{m1} + (1-p+p\cdot\eta_1)\cdot\theta_{m0} \tag{14}$$

and

$$E[\theta_m \mid q=0] = (1-p)(1-\eta_0)\cdot\theta_{m1} + (p+\eta_0(1-p))\cdot\theta_{m0} \tag{15}$$

The simplest case where the outcomes are binary and extreme (the LEA either shuts down the market, or has no impact over it) is when the LEA tries to minimize the volume of goods traded $E[2\theta_m]$, subject to a budget:

$$\min E[2\theta_m], \quad \text{subject to cost} < B \tag{16}$$

where B is a budget available to the LEA, that the LEA cannot exceed.

If the LEA observes a high quality seller then its goal is:

$$\min E[\theta_m|q=1], \quad \text{subject to } L^2\eta_1^2 < B \tag{17}$$

and if the observes a low quality seller then its goal is:

$$\min E[\theta_m|q=0], \quad \text{subject to } L^2\eta_0^2 < B \tag{18}$$

There are two cases, depending on whether or not S_H produces the high quality good. Assuming that it does, then, in order to derive the LEA's manipulation efforts η_0 and η_1 we solve the Kuhn-Tucker conditions on Equations 17 and 18. We then set $\hat{\eta}_0=\eta_0=\eta_0^*$ and $\hat{\eta}_1=\eta_1=\eta_1^*$ in order to derive η_0^* and η_1^*, the equilibrium values of η_0 and η_1 that are consistent with buyer beliefs in the Perfect Bayesian Equilibrium (the amount of review manipulation that the LEA performs coincides with the amounts expected by the consumers).

The Kuhn-Tucker conditions are:

$$\begin{cases} \frac{d\left(E[\theta_m|q=1]+\lambda_1(L^2\eta_1^2-B)\right)}{d\eta_1}=0 \\ \lambda_1(L^2\eta_1^2-B)=0 \\ \frac{d\left(E[\theta_m|q=0]+\lambda_0(L^2\eta_0^2-B)\right)}{d\eta_0}=0 \\ \lambda_1(L^2\eta_0^2-B)=0 \\ \eta_1^2<B \\ \eta_0^2<B \\ \lambda_1,\lambda_1>0 \end{cases} \tag{19}$$

It is easy to verify that the solution $\eta_0=\eta_1=\sqrt{B}/L$, $\lambda_1=\frac{p(2p-1)}{\left(\sqrt{B}/L\right)\left(1+\sqrt{B}/L\right)}$ and $\lambda_0=\frac{(1-p)(2p-1)}{\left(\sqrt{B}/L\right)\left(1+\sqrt{B}/L\right)}$ satisfies the above conditions[4]. The LEA simply uses up its entire budget in manipulating reviews, both in the case where the $q=0$ and the case where $q=1$. Note that, given how η_0 and η_1 were defined, positive values for η_0 and η_1 mean that reviews are manipulated downwards, that is, reviews are deflated.

For the solution to be consistent with the assumption that S_H produces the high quality good, it must be the case that the buyers expect S_H to produce the high quality good, and S_H must prefer the production of the high quality good to the production of the low quality good. The maximum quality production cost w' for which this is true is given by[5]: $p(1-\eta_1)\theta_{m1}^2+(1-p+p\cdot\eta_1)\theta_{m0}^2+B-w'>(1-p)(1-\eta_0)\theta_{m1}^2+(p+\eta_0(1-p))\,\theta_{m0}^2+B\Leftrightarrow w'=(\theta_{m1}^2-\theta_{m0}^2)\,(2p-1+\eta_0-p(\eta_0+\eta_1))$. For $\eta_0=\eta_1=\sqrt{B}/L$, we obtain

$$w'=\left(\theta_{m1}\left(\sqrt{B}/L,\sqrt{B}/L\right)^2-\theta_{m0}\left(\sqrt{B}/L,\sqrt{B}/L\right)^2\right)(2p-1)\left(1-\sqrt{B}/L\right) \tag{20}$$

We plot w' in Figure 3. As we can see from Equation 20, and also in Figure 3, there is always a w' below which the equilibrium will have S_H producing the high quality product, and the LEA manipulating reviews with $\eta_0=\eta_1=\sqrt{B}/L$. We will see that, depending on market parameters, the LEA may or may not be able to impact the volume of goods traded in the market, but it will always be able to influence (ex-ante) seller revenue from buyer purchases.

For $w<w'$, the impact of review manipulation to the volume of goods traded in the market is shown by the following Proposition and Corollary:

Proposition 1. *For $w<w'$ the LEA cannot affect metrics that are linear in expected product quality $x\cdot\theta+y$.*

[4] We assume that $B<L^2$, since, by our definition of η_0 and η_1, these parameters cannot exceed 1. Otherwise, the equilibrium solution should be given as $\eta_0=\eta_1=max(1,\sqrt{B}/L)$.

[5] We assume that the LEA's budget translates to seller revenue (LEA has to buy products to manipulate the market). This assumption does not affect our results, since the two B's in the two sides of the equation cancel out.

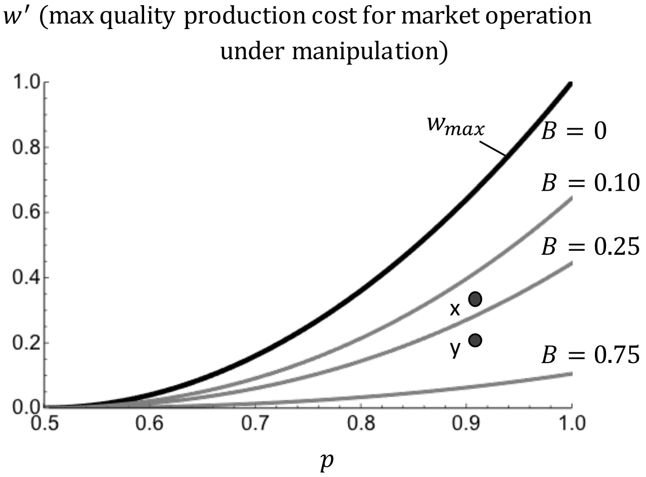

Figure 3 **max w required for market operation when the LEA manipulates reviews, for different values of the budget parameter B. In all cases $\mathbf{L = 1, f = 1/2}$**

Proof For $w < w'$ S_H produces the high quality product, and the metric with LEA manipulation is $E[x\theta_m + y] = (1-f)E[x\theta_m + y \mid q = 0] + f \cdot E[x\theta_m + y \mid q = 1]$. Using Equations 12 through 15, this sum can be shown to be constant and equal to $f \cdot x + y$, independently of η_0 and η_1. □

Corollary 1. *For $w < w'$, the LEA cannot affect the volume of products traded.*

Proof Follows immediately from Proposition 1 since volume traded is a metric linear in θ, of the form $2\theta_m$ □

The intuition why review manipulation does not reduce the volume of traded goods when $w < w'$ is that, by reducing the informativeness of the review signal, the LEA causes buyers to recognize less often high quality sellers with $q = 1$ and to instead be fooled more often with low quality sellers with $q = 0$. The decrease in the linear metric (e.g., demand) from high quality sellers is exactly offset by an increase in the metric from low quality sellers.

To summarize, for $w < w'$, in equilibrium, the high quality seller produces the high quality good, the LEA manipulates reviews with $\eta_0 = \eta_1 = \sqrt{B}/L$ and the volume of goods traded is not affected.[6]

Note that even while the LEA does not achieve any reduction in the volume of goods traded in the market, it cannot stop manipulating reviews. If it does, the buyers will continue to think that the reviews are strategically deflated, and will continue to adjust their expectations about product quality according to Equations 12 and 13, and as a result they will be overestimating product quality by (mistakenly) adjusting for LEA's expected manipulation. As a result, if the

[6] For $w < w'$ there is an alternative equilibrium where S_H produces the low quality product and buyers expect S_H to produce the low quality product. Buyers generate zero demand ($2 \cdot \theta = 2 \cdot 0 = 0$) and S_H has no incentive to produce the high quality product. No market transactions take place. We will assume for simplicity that this trivial equilibrium does not materialize, because equilibria where transactions take place, if they exist, are always favored by the market.

LEA stops manipulating reviews, product demand will increase, which is inconsistent with LEA's goal to minimize the volume of products traded. Thus, the LEA is forced to continue manipulating the market, even if it expends resources without result, simply because it is expected to.

We now turn to the case, where $w > w'$. It is easy to show that for $w > w'$ S_H will not produce the high quality good. If buyers expect S_H to produce the low quality good, they generate zero demand and S_H has no incentive to produce the high quality good. If on the other hand buyers expect S_H to produce the high quality good, by the definition of w', $\forall w$: $p(1-\eta_1)\theta_{m1}^2 + (1-p+p\cdot\eta_1)\left[\theta_{m0}\right]^2 - w < (1-p)(1-\eta_0)\theta_{m1}^2 + (p+\eta_0(1-p))\,\theta_{m0}^2$, so that S_H would prefer to produce the low quality good instead. Thus, if $w > w'$, in any PBA S_H must produce the low quality good and buyers must expect so. However, because the LEA tries to minimize the volume of goods traded, taking its budget as given, there are infinite combinations of η_0 and η_1 that are consistent with such PBAs. Obviously, if the high quality seller chooses to defect and if buyers expect this, then regardless of what η_0 and η_1 are chosen by the LEA no transactions will take place in the market.

Obviously, the most desirable equilibrium from the LEA's perspective is $\eta_0 = \eta_1 = 0$. In this equilibrium the LEA manages to shut down the DNM without expending any resources. This equilibrium has many similarities to Akkerlof's "Market for Lemons" (Akerlof 1970): there are buyers and sellers in the market who are willing to transact at a price above the product production cost, and yet no transactions take place. The nature of the equilibrium is surprising: *the LEA has achieved the complete destruction of the market, without actually manipulating any reviews*, merely by threatening review manipulation. Note that in the range $w' < w < w_{max}$ it is the presence of the LEA that causes the market to collapse: there are no other equilibria where market participants transact, even while the production cost $w < w_{max}$ is such that, in the absence of th eLEA, we would observe a functioning market, as analyzed in Section 2.1.

Arguably, when $w' < w$ the least desirable equilibrium from the LEA's perspective is the equilibrium with $\eta_0 = \eta_1 = \sqrt{B}/L$, where the LEA uses up its entire budget in manipulating reviews. Indeed, in this equilibrium the LEA expends resources even though it does not have to, since no transactions take place regardless of its actions.

However, of all possible equilibria where S_H produces the low quality product and buyers purchase no products, the equilibrium with $\eta_0 = \eta_1 = \sqrt{B}/L$ is the most robust. The reason it is the unique equilibrium in the case where even a very small amount of high quality products are available in the market. All other equilibria are unstable, in the sense that they depend on the buyers expecting exactly the minimum possible product quality. Even a small deviation from this expectation leads to the unique equilibrium with $\eta_0 = \eta_1 = \sqrt{B}/L$.

To see this consider that there is a third type of seller S_P with very low production cost w_P. For example, we can assume that this type of seller has already incurred the cost of production which are now considered sunk, so that $w_P = 0$. We will assume that there are very few such sellers in the market, and we will denote their proportion of the total population with f_P. Thus, in this model variation, the monopolist seller can be either one of three types. The seller can be of type "*High Quality*" S_H with probability f, or of type "*Low Quality*" S_L with probability $1 - f - f_P$, and of a third type denoted by S_P with probability f_P. As before, S_L can only produce at quality $q = 0$, S_H can produce at $q = 0$ at no cost, but can also produce at $q = 1$ at cost w. Now, the third type S_P can produce at $q = 1$ at no cost ($w_P = 0$).

To show how the introduction of even a small number of such sellers causes the LEA to spend its entire budget on manipulation, we will assume that f_P is strictly positive, but less than any other market parameter. In the Appendix we prove the following Proposition:

Proposition 2. *Even a very small proportion $f_P \to 0$ of sellers with $w_P = 0$ (e.g., due to sunk costs), leads to a unique equilibrium when $w > w'$ with $\eta_0 = \eta_1 = \sqrt{B}/L$*

We can now describe the Perfect Bayesian Equilibrium across three different possible outcomes:

Theorem 1. *Given* $w' = \left(\theta_{m1}\left(\sqrt{B}/L, \sqrt{B}/L\right)^2 - \theta_{m0}\left(\sqrt{B}/L, \sqrt{B}/L\right)^2\right)(2p-1)\left(1 - \sqrt{B}/L\right)$ *and* $w_{max} = (\theta_{b1}^2 - \theta_{b0}^2)(2p-1)$, *then*

Case A: *For $w_{max} < w$ no transactions take place, even in the absence of the LEA, so the LEA has no impact in the market.*

Case B: *For $w' < w < w_{max}$ the presence of the LEA causes the market to fail. S_H produces the low quality good, and product demand is zero. The LEA manipulates reviews with $\eta_0 = \eta_1 = \sqrt{B}/L$ and expends its entire budget B in the process.*

Case C: *For $w < w'$ The LEA manipulates reviews with $\eta_0 = \eta_1 = \sqrt{B}/L$ and expends its entire budget B in the process. However, S_H continues to produce the high quality good and the volume of goods traded remains unaffected.*

Proof: Follows from the preceding analysis, and Propositions 1 and 2. Case B focuses only on the most robust equilibrium, that is the unique equilibrium when even a very small proportion $f_P \to 0$ of sellers with $w_P = 0$ exists. □

Proposition 3. *Given enough budget, the LEA will cause the market to fail. Specifically, the LEA causes market failure for $B > B_{min}$, where B_{min} is the solution to $w' = w$. B_{min} increases with p, decreases with w, and is proportional to L^2.*

Proof Shown by differentiating w' (given by Equation 20) with respect to B. We can see that the derivative is always negative, and hence for large enough B we can always make $w' < w$.

Finally, it is easy to verify even with visual inspection that solving $w' = w$ for B will produce a solution that is proportional to L^2. □

2.3. Review Manipulation in Resilient DNMs

In the previous section, we saw that with large enough budget, it is always possible for the LEA to cause the DNM to fail. We also showed that with insufficient budget the volume of goods traded remains unaffected. It is natural to ask whether it is possible for a LEA whose budget does not suffice to shut down the market, to still inflict some sort of damage to the market, in pursue of a goal that is secondary to the (unattainable) reduction of market volume. Such a goal could be the reduction of seller profitability, or the reduction of the average product quality purchased.

In this Section we will study the impact of review manipulation to seller profitability and the average quality of traded products. We will assume throughout this section that the quality production cost w is low enough so that the market does not shut down due to manipulation. We must first better understand how review manipulation changes the market.

2.3.1. Impact on Quality of Traded Goods The most direct impact that review manipulation has in the market, is that it degrades the quality of the review signal and makes it harder for buyers to tell apart qualities.

The amount of information conveyed by reviews (in bits), as a function of review manipulation η_0, η_1 is given by

$$I_m(\eta_0, \eta_1) = H(Q) - H(Q \mid S) \tag{21}$$

which is the mutual information between the quality and the review signal. $H(Q)$ is the entropy of the quality signal, $H(Q) = f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{1-f}$. It equals 1 bit of information when the two seller types are equiprobable, that is, when $f = 1/2$. $H(Q \mid S)$ is the entropy of the quality, given the review signal. In other words, $H(Q \mid S)$ is the amount of uncertainty we have about quality, after we observe the reviews. Thus, overall, I_m is the amount of information that the reviews provide. From basic Information Theory (MacKay 2003):

$$H(Q \mid S) = \sum_{i,j=0,1} P(q = i; s = j) \log_2 \frac{1}{P(q = i \mid s = j)} \tag{22}$$

The amount of information (in bits) that the review signal conveys to buyers in the absence of manipulation is given by

$$I_b = I_m(0, 0) \tag{23}$$

It is easy to see that I_b is maximized for $f = 1/2$ when it equals the capacity of the binary symmetric channel with noise level p: $I_b = 1 - \left(p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}\right)$. When $f = 1/2$ and the review mechanism is perfect ($p \to 1$), it conveys the maximum amount of information possible which is 1 bit of information per channel use, as it always clarifies whether the quality level is $q = 0$ or $q = 1$, with equal probability. When the reviews are completely uninformative $p = 1/2$, Equations 21 and 23 yield zero, as expected.

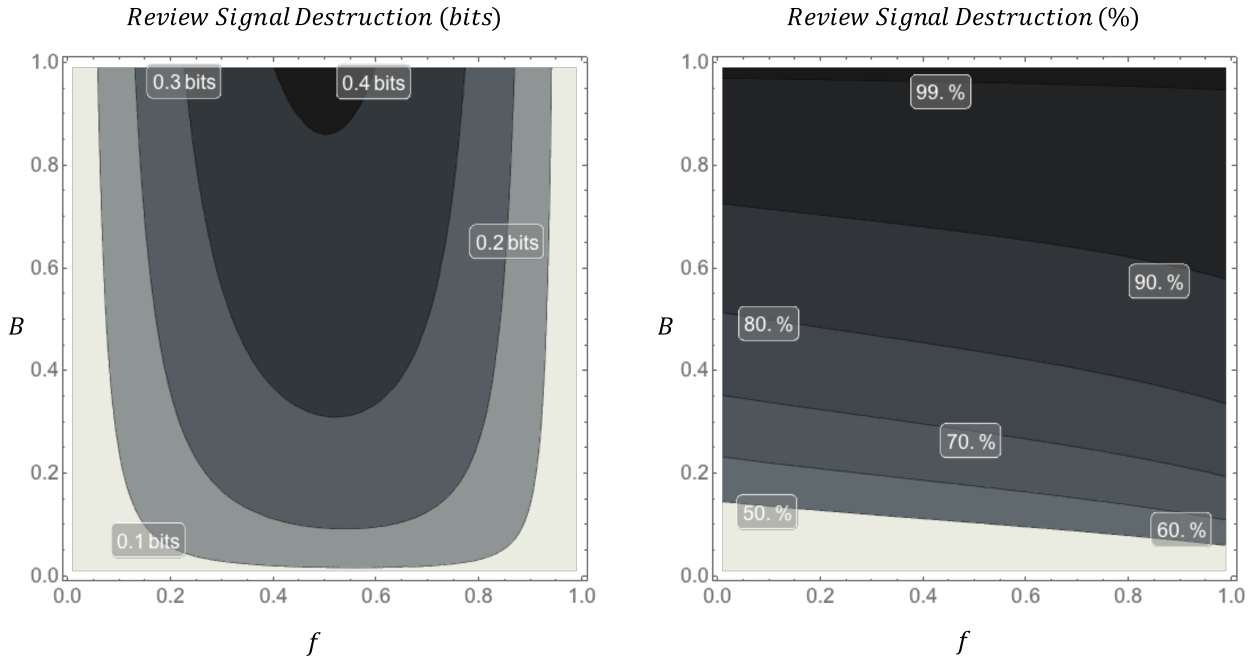

Figure 4 **Review-signal destruction measured in bits (left graph) and as a percentage of the information that reviews convey in the absence of manipulation (right graph). In both cases $p = 0.86$. w is assumed to be low enough so that the market continues to operate.**

The amount by which the LEA reduces the informativeness of the review signal when it expends its entire budget on manipulation is simply $I_b - I_m(\sqrt{B}/L, \sqrt{B}/L)$ and it is shown on the left hand side of Figure 4.

The more the LEA spends in review manipulation (higher B) the more of the review signal it destroys. Review signal destruction is more prominent when the reviews are most informative, that is when the prior f is close to 1/2. When f is close to 1 or close to zero, the reviews are not very informative, due to the strong prior. In these cases there is less of a signal for the LEA to destroy.

This is easier to see on the right hand side of Figure 4, where it is evident that the amount of review signal destroyed is roughly the same for all priors f, for a given budget level B.

Given that manipulation deteriorates the quality of the review signal, we should expect that buyers will have trouble locating the high quality product when the reviews are being manipulated. As a result, we would expect the average traded quality, that is, the quality of the average traded product, to decrease.

In the absence of manipulation, $f\%$ of the time the product is of good quality $q = 1$ and the volume sold (demand) equals $p \cdot 2\theta_{b1} + (1-p) \cdot 2\theta_{b0}$, and $(1-f)\%$ of the time the product is of bad quality $q = 0$ and the volume sold is $(1-p) \cdot 2\theta_{b1} + p \cdot 2\theta_{b0}$, so that the average quality of products sold Q_b equals $\frac{f \cdot 1 \cdot (p \cdot 2\theta_{b1} + (1-p) \cdot 2\theta_{b0}) + 0}{2f} = p \cdot \theta_{b1} + (1-p)\theta_{b0}$.

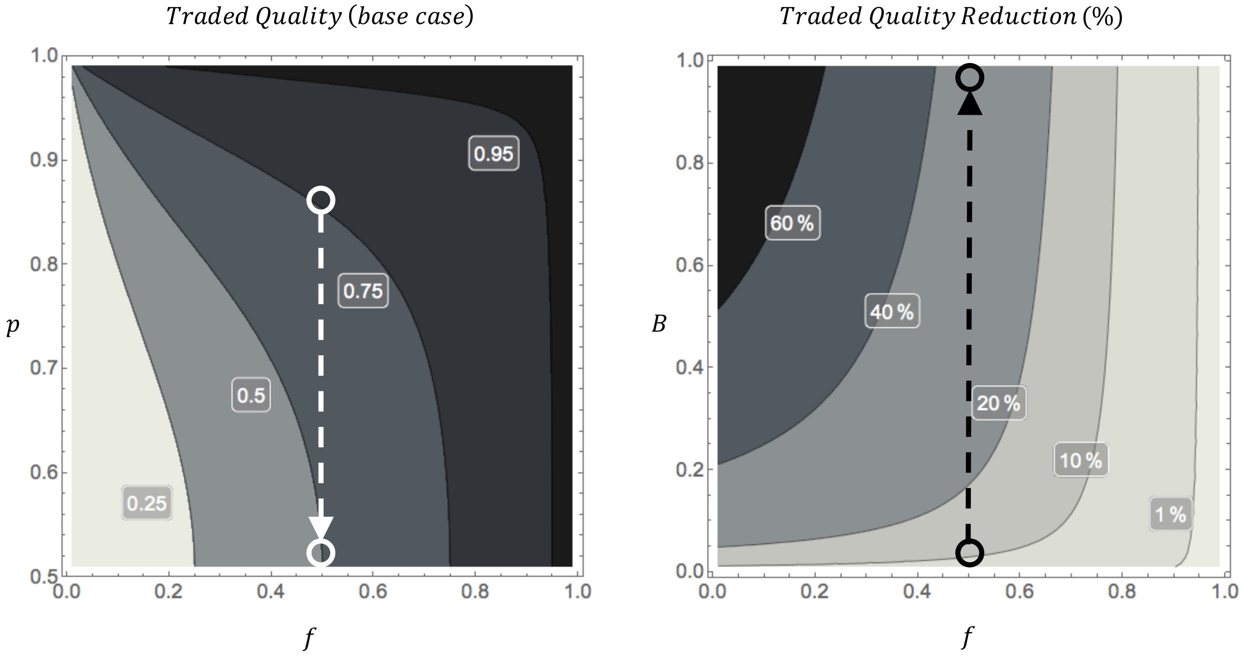

Figure 5 The quality of the average traded product in the base case (left side), and its expected percentage reduction (right side). In the latter case $\mathrm{p} = 0.86$

$$Q_b = p \cdot \theta_{b1} + (1-p)\theta_{b0} \tag{24}$$

Q_b is depicted on the left side of Figure 5. As expected, traded quality increases with the prior f. Note how an efficient review mechanism (high p) can guarantee high traded quality, even with a low prior f, that is, even when the seller is very likely of the Low Quality type. With a highly accurate review signal, the buyers are able to avoid low quality products and purchase the high quality products instead.

When the LEA manipulates the reviews and makes them less informative, more products of low quality are being sold at the expense of high quality products, since demand remains the same as per Proposition 1. Specifically when the LEA manipulates reviews the average quality of product sold reduces to

$$Q_m(\eta_0, \eta_1) = p \cdot (1-\eta_1)\theta_{m1} + (1 - p + p \cdot \eta_1)\theta_{m0} \tag{25}$$

The percentage reduction of average quality traded $\frac{Q_b - Q_m(\sqrt{B}/L, \sqrt{B}/L)}{Q_b}$ is depicted on the right hand side of Figure 5. As expected, the more intense the manipulation effort, the greater the percentage reduction in traded quality. For example, note the black arrow on the right hand side, which depicts the change in traded quality by moving from a zero budget investment in manipulation to a unit budget investment in manipulation. As per Figure 4, this will completely destroy the informativeness of the reviews, and it would be equivalent to reducing the informativeness of the reviews in the base case from their initial value of $p = 0.86$ to $p = 1/2$, (white arrow on the left hand side). This makes the reviews completely uninformative. As we can see on the left side, this would reduce average quality traded from 0.75 units to 0.50, or a 33% reduction, consistent with what the right hand side shows us. Using the same reasoning, it is easy to see that manipulation does not have much impact on average traded quality when the prior f is very high. In that case, as we can also verify from the left hand side of the Figure, the quality the reaches buyers is almost always high, regardless of the efficiency of the review mechanism.

2.3.2. Impact on Seller Profitability We have seen that, even when the LEA does not have sufficient budget to shut down the market, it still manages to reduce the quality of traded goods, via its detrimental impact on the review signal. However, this implies that the LEA favors the low quality seller at the expense of the high quality seller: S_L sells more to buyers and S_H sells less. Because revenue is convex in expected quality (e.g., see Equation 6), in aggregate, manipulation should decrease seller revenue from buyer purchases.

In this section, we assume that a LEA whose budget is not sufficient to shut down the market, aims instead to reduce the average seller profits. Now, the mechanism by which the review mechanism is being manipulated becomes important. If we assume that LEA's manipulation budget is transfered to sellers via LEA's product purchases, then we must also account for the funds that the LEA injects into the market. In that case, we find that LEA manipulation may lead to increased seller profits.

In the absence of manipulation, f% of the time the product is of good quality $q = 1$ and the seller earns $p \cdot \theta_{b1}^2 + (1-p) \cdot \theta_{b0}^2$, and $(1-f)$% of the time the product is of bad quality $q = 0$ and the seller earns $(1-p) \cdot \theta_{b1}^2 + p \cdot \theta_{b0}^2$, so that seller expected revenue R_b equals

$$R_b = f\left(p \cdot \theta_{b1}^2 + (1-p)\theta_{b0}^2\right) + (1-f)\left((1-p)\theta_{b1}^2 + p \cdot \theta_{b0}^2\right) \tag{26}$$

Similarly, when the LEA manipulates the reviews, in equilibrium, sellers' ex-ante revenue is

$$R_m(\eta_0, \eta_1) = f\left(p(1-\eta_1)\theta_{b1}^2 + (1-p+p\cdot\eta_1)\theta_{b0}^2\right) + (1-f)\left((1-p)(1-\eta_0)\theta_{b1}^2 + (p+\eta_0(1-p))\,\theta_{b0}^2\right) + B \tag{27}$$

Assuming that the LEA expends its entire budget in manipulating reviews, then the percentage change of seller revenue $\frac{R_b - R_m(\sqrt{B}/L, \sqrt{B}/L)}{R_b}$ is depicted in Figure 6. We plot the change in seller revenue as a function of the budget B and the manipulation cost parameter L.

Two features stand out in the plot. First, the LEA's task becomes more and more difficult, as we increase L: at all budget levels, the LEA is worse off if we increase the manipulation cost. Second, the LEA may actually increase seller revenue, if the budget B exceeds the *organic* decrease in seller

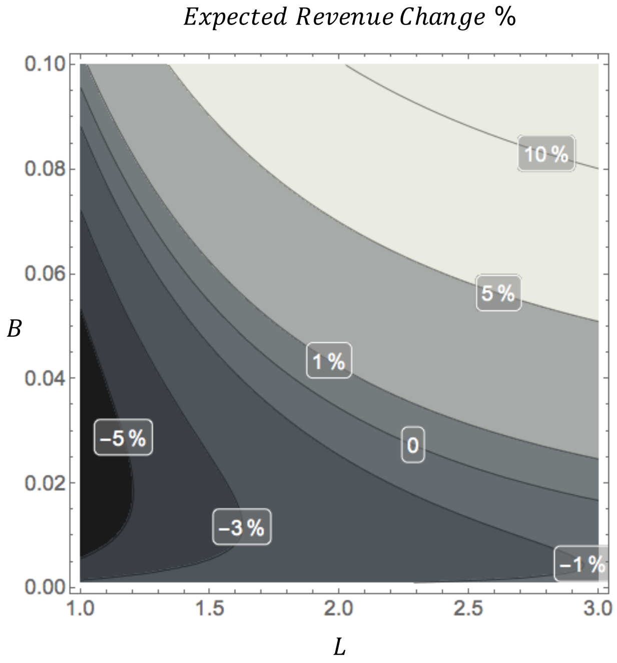

Figure 6 Expected change (percentage) in seller revenue, due to the operation of the LEA, accounting for the increased seller revenue due to the LEA's operation. Parameter values used: $\mathbf{f = 1/2}$, $\mathbf{p = 0.95}$

revenue due to buyer purchases. In contrast to L, the impact of B on the change in seller revenue is non-monotonic. In the market instance of $f = 1/2$ and $p = 0.95$, the maximum revenue decrease that the LEA can achieve is about 7% for $B \approx 0.025$, when $L = 1$.

Figure 7 summarizes the four different regimes that we have observed in our analysis. As per Theorem 1, for $w > w_{max}$ (see Equation 11) only the pooling equilibrium is possible and no transactions take place. For $w' < w < w_{max}$ (see Equation 20), it is review manipulation by the LEA that gives rise to the pooling equilibrium and causes the market to fail. For $w < w'$ the market is resilient in review manipulation and the volume of goods traded remains unchanged. In the $w < w'$ region, average traded quality reduces, but depending on the values of the parameters that describe the market, average seller revenue may decrease or increase. For $f = 1/2, L = 1$, and $B = 0.05$, p must be larger than $p' \approx 0.87$ for the LEA to be able to reduce average seller revenue.

Ideally, the LEA would like to use as large a budget as possible: as we increase B, w' reduces and the size of the dark gray area increases; that is, the LEA is able to cause market failure for a wider range of market parameters. But the additional budget required for the LEA to have an impact, may not be available to it. It then becomes important for the LEA to use an optimal manipulation strategy. This is the topic of the next Section.

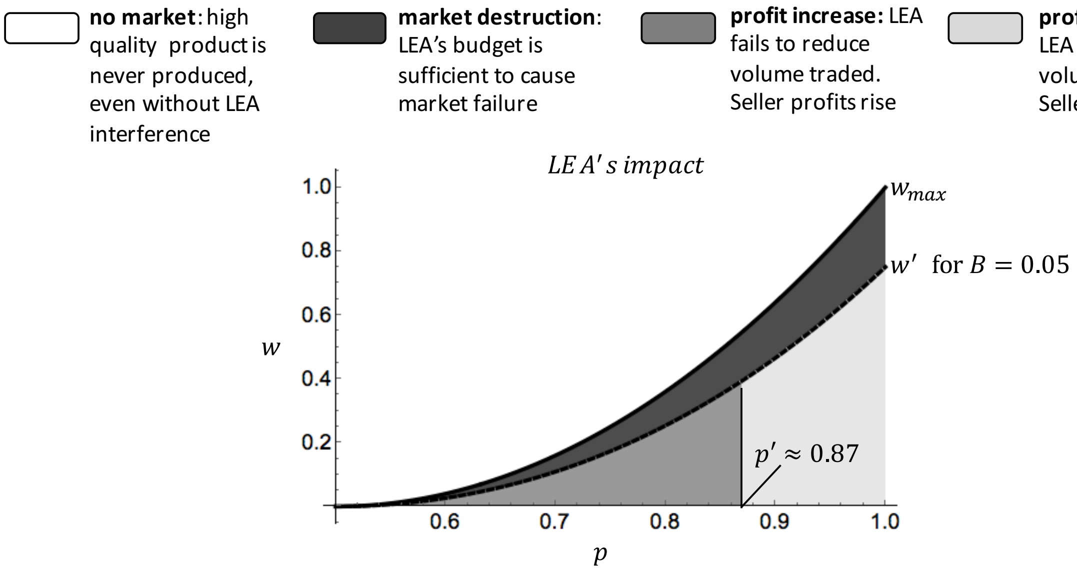

Figure 7 **Summary of the different observable market regimes. Parameter values used: $\mathbf{f = 1/2}$, $\mathbf{L = 1}$, $\mathbf{B = 0.05}$**

3. Efficient Review Manipulation

This Section explores the efficiency of the LEA's operations. We will introduce a *utility-based approach* that is optimal in the sense that it can cause market failure at a minimum cost. The utility that we assume that the LEA is trying to maximize is given by:

$$U = E[\theta_b^2] - E[\theta_m^2] - (\alpha + 1)L^2\eta^2 \tag{28}$$

where η is the LEA's manipulation effort and $\alpha \in (-1, \infty)$ is a free parameter.

One way to interpret parameter α is to view it as a measure of *aggressiveness* that the policy maker endows the LEA with. The lower α is, the more aggressive the LEA will be in manipulating the reviews, and the more resources it will expend doing so.

The intuition for the chosen form of Equation 28 is as follows. A LEA that tries to damage a DNM in an efficient manner, would need a way to gauge its impact versus the marginal resources expended. It could for example continue to invest in market manipulation, as long as any extra *dollar* put into manipulating the market reduces seller profitability by α dollars or more. The LEA would not manipulate reviews beyond the limit where each additional dollar spent does not achieve this goal. A utility function that describes this behavior is given by $\frac{E[\theta_b^2] - E[\theta_m^2] - L^2\eta^2}{\alpha} - L^2\eta^2$. By scaling by a factor α we derive Equation 28.

Equation 28 can also be viewed as a *strategy generator*, because the policy maker, by choosing different values for the α parameters, can endow the LEA with different strategies. A few examples are as follows. For $\alpha >> 1$ the LEA acts conservatively, only manipulating when it knows that it will reduce seller profitability many times more than the resources it expended. For $\alpha = 1$ the

LEA is prepared to keep manipulating as long as its marginal contribution to the reduction of seller revenue exceeds its marginal manipulation cost. For $\alpha = 0$ the LEA disregards its own cost, valuing only the reduction in seller profit. This is equivalent to a budget-based approach, where the LEA is asked to minimize seller cost, considering its budget as sunk. Finally, for $\alpha = -1$ the LEA seeks to destroy the review signal completely, without considering the impact to its cost, or to seller profitability: for $\alpha = -1$ seller profitability will most likely increase, due to the high cost of the large manipulation effort.

Depending on whether or not the LEA observes a high or a low quality seller, it will try to maximize either one of the following two equations:

$$U_1(\eta_1) = E[\theta_b^2|q=1] - E[\theta_m^2|q=1] - (\alpha+1)L^2\eta_1^2 \tag{29}$$

$$U_0(\eta_0) = E[\theta_b^2|q=0] - E[\theta_m^2|q=0] - (\alpha+1)L^2\eta_0^2 \tag{30}$$

In order to derive the LEA's equilibrium manipulation efforts η_0 and η_1, we take the appropriate F.O.C. on Equations 29 and 30. We then set $\hat{\eta_0} = \eta_0 = \eta_0^*$ and $\hat{\eta_1} = \eta_1 = \eta_1^*$ in order to derive η_0^* and η_1^*, the equilibrium values of η_0 and η_1 that are consistent with buyer beliefs in the Perfect Bayesian Equilibrium. In other words, the amount of bad reviews that the LEA uses coincides with the amounts expected by the consumers.

It is easy to see that equilibrium η_0^* and η_1^* decrease with α. Based on our discussion above, and depending on how much the policy maker penalizes or encourages LEA manipulation through the factor α, the policy maker can match the cost of any other given strategy. The expected cost of LEA's strategy $f\eta_1^{*2} + (1-f)\eta_0^{*2}$ can be set to arbitrarily close to zero for very high α, all the way to a maximum cost required to completely destroy the review signal, for $\alpha = -1$. Another way to see this is that, by starting with a very large α and reducing it down to $\alpha = -1$, we achieve review informativeness (Equation 21) that ranges from a high of I_b (review informativeness in the base case; see Equation 23) all the way down to zero. When $\alpha = -1$ Equations 29 and 30 yield $\eta_0^* = 1-p$ and $\eta_1^* = p$. This implies, that regardless of whether $q = 0$ or $q = 1$, and independently of f, the probability that signal $s = 1$ will be observed will always be $p(1-p)$. The review mechanism is completely uninformative, as can be verified by setting $\eta_0^* = 1-p$ and $\eta_1^* = p$ in Equation 21.

We refrain from showing the closed form solutions for η_0^* and η_1^* due to their complexity. The important observation is that they can be seen to be unique and positive. This means that the LEA will *not* try to make the low quality sellers with $q = 0$ look good by giving them positive reviews, while giving bad reviews to high quality sellers with $q = 1$. Instead the LEA will only give negative reviews, but more so for the high quality sellers.

We depict η_0^* and η_1^* in Figure 8, for $f = 1/2$ and $\alpha = 2$. The LEA manipulates reviews as long as any extra dollar put into manipulating the market, reduces expected seller profitability by two

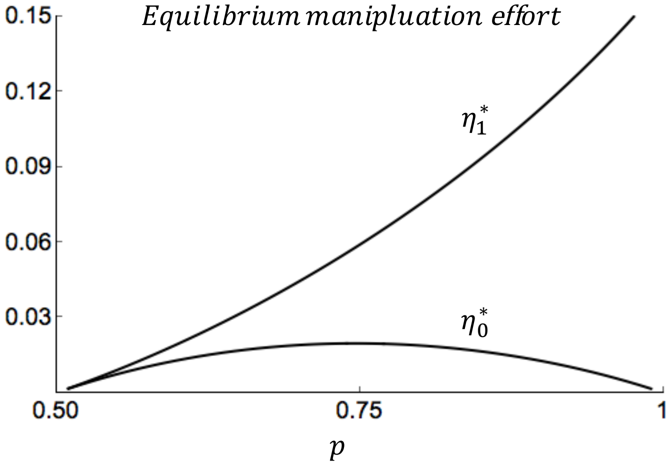

Figure 8 LEA manipulation effort in equilibrium, when the LEA maximizes its utility, given by Equation 28. Parameter values used: $\mathbf{f = 1/2}$, $\alpha = \mathbf{2}$

dollars, or more. It is immediately obvious that the LEA is more efficient, compared to the budget-driven approach of Section 2, where the LEA always spent its entire budget, regardless of whether $q = 0$ or $q = 1$. A utility-driven LEA that tries to maximize Equations 29 and 30 will not manipulate the market for any $\alpha > -1$ when the review mechanism does not convey information to the buyers ($p = 1/2$). The LEA knows that it cannot reduce the informativeness of the mechanism any further and will set ($\eta_0^* = \eta_1^* = 0$) to reduce manipulation cost. The LEA will also not try to manipulate reviews when $q = 0$ and the review mechanism conveys perfect quality information ($p = 1$). In that case the LEA knows that the perfect review signal will always reveal the poor quality to the buyers who will refrain from purchasing the product.

This efficient use of resources on the part of LEA drives the following Proposition:

Proposition 4. *With a utility-driven LEA that maximizes Equation 28, and assuming that $\alpha > -1$, even a very small proportion $f_P \to 0$ of sellers with $w_P = 0$ (e.g., due to sunk costs), leads to a unique equilibrium when $w > w'$ with $\eta_0 = \eta_1 = 0$*

Proof: see Appendix

Proposition 4 shows that the only robust equilibrium when the LEA shuts down the market with the utility-driven strategy is the equilibrium where the LEA expends no resources. In other words, it manages to cause market failure merely by threatening to manipulate the market. The intuition, is that, in contrast to the budget-based approach, the LEA does not consider its resources as sunk (assuming $\alpha \neq -1$) so that it does not overspend when only a small number of high quality products is available in the market.

Theorem 2. *Assume that an arbitrary review manipulation strategy S yields equilibrium manipulation efforts η_{1s} and η_{0s}, and causes the market to fail. Then, we can always choose an α that* *yields a manipulation strategy S^{α} based on the utility based approach of equation $E[\theta_b^2] - E[\theta_m^2] - (\alpha+1)L^2\eta^2$), that is at least as efficient as S. Specifically, S^{α} also achieves market failure at an equal or lower total expected cost.*

Proof: If there exists a strategy S that achieves market failure, then there must also exist a strategy S^{α} that also achieves this. This is because market failure is caused when the informativeness of the review mechanism $I_m(\eta_0, \eta_1)$ is low enough for S_H to prefer to produce the low quality good, leading to the pooling equilibrium where only low quality is available. We saw that by manipulating α we can achieve all allowed values for I_m allowed in the market, monotonically from $I_m = I_b$ for $\alpha \to \infty$, all the way down to $I_m = 0$ for $\alpha = -1$. Thus, if there is value for I_m achievable by strategy S that leads to the pooling equilibrium, then there will also be an α so that the utility-based approach also causes market failure. From Proposition 4, we know that the LEA following strategy S^{α} expends zero resources when shutting down the market, so that no other strategy can do better.

□

4. Policy Recommendations

Proposition 3 and Theorem 1 have important policy implications. The policy maker who controls the LEA's budget can always provide the LEA with enough resources to render a DNM inoperable. However, it may not be practical to do so, as the required budget may be very high. The three main factors that can increase the budget required to successfully attack a DNM are the a-priori effectiveness of the review mechanism p, the quality production cost w, and the manipulation cost parameter L.

In general, the more effective the review mechanism is (higher p), the higher B_{min}. That is, a higher p makes it more costly for the LEA to manipulate the review mechanism so that the high quality seller can no longer depend on it to deliver high quality, and instead prefers to defect to producing low quality. Similarly, the lower the quality production cost w (other things equal) the more review manipulation is needed to make the production of the high product unattractive.

However, what is arguably the most important hurdle in attacking DNMs through the review mechanism, is the fact that $B_{min} \propto L^2$. That is because the market operator can influence L through countermeasures that are designed to increase review manipulation cost. Indeed, as we discussed, one way to interpret the parameter L is that it represents the number of truthful reviews needed before a fake review can be safely used without exposing the LEA as a strategic manipulator. The stricter the market operator chooses to be with reviewers that substantially deviate from the mean, the higher L becomes, and the higher the budget required to successfully attack the DNM. However, actions on the part of the DNM to raise L may carry a significant cost for the market, as they are likely to offer cover for sellers to strategically cheat buyers. For example, a rule that flags a buyer as potentially LEA when that buyer significantly disagrees with the mean review once in every L transactions, would allow a seller to systematically cheat. For example, by cheating with impunity in transactions with buyers that have not yet accumulated L posted reviews. The interesting dynamics of DNMs countermeasures and the costs that they impose on market operation are beyond the scope of our paper, where we accept that the functional form of the manipulation cost subsumes any counter-manipulation efforts on the part of the DNM operator.

Another interesting policy implication is that, if the goal of the policy maker who controls the LEA budget is the reduction of the volume of goods traded in the DNM, then the provision of insufficient funds for the manipulation effort does not achieve partial results. As per Corollary 1 it achieves nothing at all. For example, if the LEA overestimates w and/or underestimates L and p it will expend its entire budget without reducing the volume of traded goods. For example, in Figure 3, if the policy maker believes that the market operates at point x it may try to commit budget $B = 0.25$ to the LEA. However, if that assessment proves incorrect and the market really operates at point y then the entire budget of $B = 0.25$ will be spent without reducing the volume of goods traded.

In fact, we showed that the situation can be even worse. When the LEA does not succeed in reducing the volume of goods traded, two other undesirable outcomes may occur. First, as we saw in Section 2.3.1 the LEA's detrimental affect on the review signal means that the average quality of goods decreases. In certain cases, e.g., for illegal drugs, this may mean that the LEA puts the health of buyers at greater risk. Second, as we saw in Section 2.3.2, seller profits may actually increase (depending on market parameters) as the LEA injects funds into the DNM. Clearly, under these conditions, LEA manipulation is socially harmful and should not be attempted at all.

Overall, our findings show that review manipulation in DNMs is a highly promising method that LEAs can use to check these markets' proliferation and growing popularity. However, our findings also underline the importance for the LEA to have accurate information on the cost structure of market participants, on the effectiveness of the review mechanism, as well as on the details and the effectiveness of potential review manipulation countermeasures, set by the market operator. Only then can the LEA confidently attempt to reduce the volume of goods traded, or failing that, to make these DNMS less profitable for the sellers who operate there.

5. 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 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.

We employed a simple model of a DNM with two levels of quality and a simple review mechanism qualities. Our main findings can be summarized as follows. We first showed that, given sufficient resources, the LEA can decrease the informativeness of the review mechanism to the point where the market ends up in a pooling equilibrium of low quality products. Effectively, the LEA manages to cause the market to collapse. However, if the LEA does not have sufficient resources to bring the DNM to such a tipping point, it expends its resources without affecting the volume of goods traded in the market. In that case, the LEA may be better off pursuing secondary goals, such as the reduction of seller profitability. We proceeded to show that, indeed, even if the LEA cannot reduce the volume of products traded in the market, it may still be able to make the DNM less profitable for its sellers. However, this is not guaranteed: for certain ranges of market parameters, and also accounting for the funds that the LEA injects into the DNM, the LEA can even cause a net increase in seller profitability. Finally, we were able to derive the optimal LEA strategy that achieves market failure

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. For example, LEA strategies that aim to reduce the effectiveness of the review mechanism by randomizing review ratings, or targetted strategies where the LEA chooses ex-ante which sellers to attack, may in fact prove more realistic and/or more effective ways for LEAs to achieve their goals.

Appendix A: Key symbols used in the notation in the Binary Model

Key Symbol	Definition
Decision Variables	
r	Product price
η_0, η_1	amount by which the LEA manipulates the reviews of a product with quality zero, and one, respectively
Model Parameters	
q	Product quality (either zero, or one)
f	Probability that the seller will be of type "*High Quality*"
w	Production cost for high quality product
s	Observable review signal: $s = 1$ is the "good reviews" signal, and $s = 0$ is the "bad reviews" signal
p	Probability that the review signal will match product quality in the absence of manipulation (accuracy-informativeness of the review mechanism). $p = 1/2$ for uninformative reviews, $p = 1$ for reviews that convey perfect information
L	unit cost of manipulation (total cost is $c(\eta) = L^2 \cdot \eta^2$)
B	LEA's budget available for review manipulation
Derived values	
D	Product demand. D_b and D_m with and without LEA review manipulation
R	Seller revenue. R_b and R_m with and without LEA review manipulation
U	LEA utility. U_{volume} and U_{value} depending on goals
θ	Buyers' expected quality after they observe s: $\theta = E(q \mid s)$. Specifically $\theta_0 = E(q \mid s = 0)$, and $\theta_1 = E(q \mid s = 1)$. Also θ_b, θ_m are θ with and without LEA manipulation, respectively. Thus, for example, θ_{b0} is $E(q \mid s = 0)$ without LEA manipulation
Q	Quality of the average traded product. Q_b in the base case, without LEA manipulation and Q_m with LEA manipulation
α	LEA would be willing to spend 1 dollar in order to reduce Seller revenue by $\alpha > 1$ dollars (when it targets the value of trade)
β	LEA would be willing to spend 1 dollar in order to reduce the volume of goods that reach buyers by β units (when it targets the volume of trade)
I	Information conveyed by the reviews, in bits. I_b and I_m with and without LEA review manipulation,

Appendix B: Proofs and derivations

Proof of Proposition 2 *Even a very small proportion $f_P \to 0$ of sellers with $w_P = 0$ (e.g., due to sunk costs), leads to a unique equilibrium when $w > w'$ with $\eta_0 = \eta_1 = \sqrt{B}/L$*

Proof: When S_H is expected to produce at quality $q = 0$ and $f_p > 0$, we have that:

$$\theta_{m_1} = \frac{p(1-\hat{\eta}_1)f_p}{(1-p)(1-\hat{\eta}_0)(1-f_p)+p(1-\hat{\eta}_1)f_p} \text{ and } \theta_{m_0} = \frac{((1-p)(1-\hat{\eta}_1)+\hat{\eta}_1)f_p}{(p(1-\hat{\eta}_0)+\hat{\eta}_0)(1-f_p)+((1-p)(1-\hat{\eta}_1)+\hat{\eta}_1)f_p}.$$

We will argue, that when $f_p \in (0,1/2)$, in equilibrium, it must be the case that $\theta_{m_1} > \theta_{m_0}$. This is so because if $\theta_{m_1} = \theta_{m_0}$, then $E[\theta_m|q=1]$ is invariant in η_1 and $E[\theta_m|q=0]$ is invariant in η_0, and hence LEA prefers to set $\eta_1 = \eta_0 = 0$. But if this expected by consumers (as it must be the case in equilibrium), we should have $\hat{\eta}_1 = \hat{\eta}_0 = 0$ which trivially induces $\theta_{m_1} > \theta_{m_0}$ for every admissible parameter values. Moreover, if $\theta_{m_1} < \theta_{m_0}$, then $E[\theta_m|q=1]$ is strictly decreasing in η_1 and $E[\theta_m|q=0]$ is strictly decreasing in η_0, and hence LEA prefers to set $\eta_1 = \eta_0 = -\sqrt{B}/L$. If consumers expect $\hat{\eta}_1 = \hat{\eta}_0$, then $\theta_{m_1} < \theta_{m_0}$ if and only if:

$$\hat{\eta}_0 < \frac{-f_p-p+2f_pp}{1-f_p-p+2f_pp} < -1.$$

for every $f_p \in (0,1/2)$ and every $p \in (1/2,1)$. That is, when $f_p \in (0,1/2)$, $\theta_{m_1} < \theta_{m_0}$ is sustainable in equilibrium only if $\hat{\eta}_1 = \hat{\eta}_0 = -\sqrt{B}/L < -1$ which is against our assumptions on the admissible values of B and L. Finally, if $\theta_{m_1} > \theta_{m_0}$, then $E[\theta_m|q=1]$ is strictly increasing in η_1 and $E[\theta_m|q=0]$ is strictly increasing in η_0, and hence LEA prefers to set $\eta_1 = \eta_0 = \sqrt{B}/L$. Since, this expected by consumers (as it must be the case in equilibrium), we should have $\hat{\eta}_1 = \hat{\eta}_0 > 0$ which correctly induces $\theta_{m_1} > \theta_{m_0}$ for every admissible parameter values. In specific, in equilibrium we must have $\hat{\eta}_1 = \hat{\eta}_0 = \sqrt{B}/L$ and hence:

$$\theta_{m_1} = \frac{p(1-\sqrt{B}/L)f_p}{(1-p)(1-\sqrt{B}/L)(1-f_p)+p(1-\sqrt{B}/L)f_p} \text{ and } \theta_{m_0} = \frac{((1-p)(1-\sqrt{B}/L)+\sqrt{B}/L)f_p}{(p(1-\sqrt{B}/L)+\sqrt{B}/L)(1-f_p)+((1-p)(1-\sqrt{B}/L)+\sqrt{B}/L)f_p}.$$

This is true for every $f_p \in (0,1/2)$ and hence for any arbitrarily small positive value of f_p. This suffices to establish validity of the proposition.

[Notice moreover that $\lim_{f_p\to 0}\theta_{m_1} = \lim_{f_p\to 0}\theta_{m_0} = 0$ and that $\lim_{f_p\to 0}\frac{\theta_{m_1}}{\theta_{m_0}} = \frac{p(p(1-\sqrt{B}/L)+\sqrt{B}/L)}{(1-p)(1-p(1-\sqrt{B}/L))} > 0$]

□

Proof of Proposition 4 *With a utility-driven LEA that maximizes Equation 28, and assuming that* $\alpha > -1$*, even a very small proportion* $f_P \to 0$ *of sellers with* $w_P = 0$*, leads to a unique equilibrium when* $w > w'$ *with* $\eta_0 = \eta_1 = 0$

Proof: Similarly to the proof of proposition 2 we consider here as well that S_H is expected to produce at quality $q=0$ and $f_p > 0$. In an equilibrium of such a case, it must be the case that:

$$\frac{\partial U_1}{\partial \eta_1}|_{\eta_1=\hat{\eta}_1} = 0 \Leftrightarrow -\frac{\partial E[\theta_m^2|q=1]}{\partial \eta_1}|_{\eta_1=\hat{\eta}_1} - 2(\alpha+1)L^2\hat{\eta}_1 = 0.$$

Notice that:

$$\lim_{f_p\to 0}\frac{\partial E[\theta_m^2|q=1]}{\partial \eta_1} = \lim_{f_p\to 0}[-\frac{f_p^2p^3(-1+\hat{\eta}_1)^2}{(-1+p+\hat{\eta}_0-p\hat{\eta}_0+f_p(1-\hat{\eta}_0+p(-2+\hat{\eta}_0+\hat{\eta}_1)))^2} + \frac{f_p^2p(1+p(-1+\hat{\eta}_1))^2}{(p+\hat{\eta}_0-p\hat{\eta}_0+f_p(1-\hat{\eta}_0+p(-2+\hat{\eta}_0+\hat{\eta}_1)))^2}] = 0$$

for every reasonable parameter values and beliefs.[7] Hence, $\frac{\partial U_1}{\partial \eta_1}|_{\eta_1=\hat{\eta}_1} = 0$ suggests that $\hat{\eta}_1 \to 0$ when $f_p \to 0$. Similarly, we can show that $\hat{\eta}_0 \to 0$ when $f_p \to 0$, which suffices to establish validity of the proposition.

□

References

Akerlof, George A. 1970. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics* **84**(3) 488–500.

Ball, James, Bruce Schneier, Glenn Greenwald. 2013. NSA and GCHQ target tor network that protects anonymity of web users. The Guardian. URL `http://www.theguardian.com/world/2013/oct/04/nsa-gchq-attack-tor-network-encryption`.

[7] Notice that when all $\hat{\eta}_0$, $\hat{\eta}_1$ and f_p converge to zero, we have that $\frac{\partial E[\theta_m^2|q=1]}{\partial \eta_1}$ converges to zero for every $p \in (1/2,1)$.

Christin, N. 2013. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. *Proceedings of the 22nd International World Wide Web Conference (WWW'13)*. 213–224.

Dellarocas, Chrysanthos. 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science* **52**(10) 1577–1593.

Dingledine, R., N. Mathewson, P. Syverson. 2004. Tor: The second-generation onion router. *Proceedings of the 13th USENIX Security Symposium*. 303–320.

Economist. 2014. The amazons of the dark net. The Economist, November 1st 2014, Print Edition.

Fowler, Geoffrey A. 2012. Tor: An anonymous, and controversial, way to web-surf. The Wall Street Journal. URL `http://online.wsj.com/article/SB10001424127887324677204578185382377144280.html`.

Greenberg, A. 2014a. Drug market agora replaces the silk road as king of the dark net. Wired. URL `http://www.wired.com/2014/09/agora-bigger-than-silk-road`.

Greenberg, A. 2014b. FBI story of finding silk roads server sounds a lot like hacking. Wired. URL `http://www.wired.com/2014/09/fbi-silk-road-hacking-question/`.

He, Stephen X., Samuel D. Bond. 2015. Why is the crowd divided? attribution for dispersion in online word of mouth. *Journal of Consumer Research* **41**(6) 1509–1527.

Kroll, Joshua A, Ian C Davey, Edward W Felten. 2013. The economics of bitcoin mining, or bitcoin in the presence of adversaries. *The Twelfth Workshop on the Economics of Information Security (WEIS 2013)*.

MacKay, David J.C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Nark, Jason. 2013. The repentant informant. Philly.com. URL `http://articles.philly.com/2013-08-27/news/41459372_1_drug-war-informant-prescription-drugs`.

Reed, M. G., P. F. Sylverson, D. M. Goldschlag. 1998. Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications* **16**(4) 482–494.

Reid, Fergal, Martin Harrigan. 2012. *An Analysis of Anonymity in the Bitcoin System*. Springer, 197–223.

Soska, K., N. Christin. 2015. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. *Proceedings of the 24th USENIX Security Symposium*.