Routledge
Taylor & Francis Group

Cooperation and distrust in extra-legal networks: a research note on the experimental study of marketplace disruption

Lonie Sebagh[a], Jonathan Lusthaus[a], Edoardo Gallo[b], Federico Varese[a] and Sean Sirur[c]

[a]Department of Sociology, University of Oxford, Oxford, UK; [b]Department of Economics, University of Cambridge, Cambridge, UK; [c]Department of Computer Science, University of Oxford, Oxford, UK

ABSTRACT

Cybercriminal markets serve as hubs for offenders and enable the sale of illegal goods and services. Thus far, the primary tactics that have been employed against these sites are arrests of cybercriminals and takedowns of marketplace infrastructure. This research note examines a different genus of disruptive strategy: attacks on user reputation. In this area, there has been some scholarly discussion of slander and Sybil operations as a means of fostering distrust. But carrying out empirical work on the effectiveness of these tactics is challenging. This research note presents a possible method for investigating this topic: social laboratory experiments. It reports on a feasibility pilot study inspired by cybercrime disruption, but which speaks to a broader range of extra-legal markets.

ARTICLE HISTORY
Received 18 Jan 2021
Accepted 13 Jan 2022

KEYWORDS
Cybercrime; extra-legal marketplaces; disruption; experiments; reputation; trust

Introduction

Cybercrime is a major burden for businesses and individuals. Breaches happen with great regularity because data can be monetised, as just like in any other industry – legal or illicit – trading complements production. This process requires cybercriminal markets, which serve as hubs for offenders and enable the sale of illegal goods and services[1]. These marketplaces are online, have some advantages over offline markets, and have attained increased visibility over the last two decades[2]. They have become variously associated with terms like the 'Dark Web' and 'Darknet'.

By disrupting these marketplaces, the burden of cybercrime may be reduced. Disruption is already a stated aim of some law enforcement agencies[3]. But, thus far, the primary methods that have been employed are arrests of cybercriminals and takedowns of marketplace infrastructure, which remove the sites themselves[4]. While further research is warranted in this space, some initial work has questioned the value and impact of these conventional tactics. These studies have suggested that the cybercriminal trading is largely displaced temporally and/or spatially[5].

This research note examines a different genus of disruptive strategy: attacks on user reputation. Alongside the 'hammer' of arrests and takedowns, are softer approaches that are aimed more at pulling economic levers to manipulate markets. While one might wish for complete market failure, damaging the efficiency of such markets may also hold some

CONTACT Lusthaus, Jonathan jonathan.lusthaus@sociology.ox.ac.uk Department of Sociology, University of Oxford, UK

value. In this area, there has been some scholarly discussion of slander and Sybil operations. These tactics are aimed at fostering distrust between marketplace users by respectively leaving false feedback on vendors and increasing the number of defaults on trades[6]. While these tactics have been discussed at a conceptual level, little empirical research has been undertaken about their effectiveness.

This research note contributes to addressing this gap. As will be discussed in greater detail below, there are a number of practical and ethical challenges to intervening in cybercrime markets in the 'wild' as part of an academic study. It is also difficult to gain data on law enforcement or industry operations of this kind (assuming they exist). As a result, we have chosen to employ a laboratory experiment to study this topic. This approach is a standard methodology in economics and other social sciences[7]. Yet its application to cybercrime would be novel. Thus far, only limited use of experiments has been employed in cybercrime research by social scientists, and this has been restricted to field experiments[8].

While cybercrime disruption inspires this experiment, we do not (and cannot) perfectly replicate the nature of this world and its offenders within a laboratory setting (for further details see the Discussion section). Instead, the study's contribution is in identifying and measuring disruption within extra-legal marketplaces, as part of a broader social process. We seek to address two questions: 1) How do traders respond to Sybil and slander operations in unregulated markets with no face-to-face interaction?; 2) Is one type of operation more effective than the other? Addressing these more foundational questions has direct relevance to cybercriminals and their online marketplaces, as but one well-known form of unregulated market. While the design of this experiment touches on broader economic principles, it does ensure that fundamental components of cybercrime marketplaces and their disruption are still present.

This research note is divided in five parts: the first part will provide readers with some background on cybercrime marketplaces; the second component will examine relevant theory on reputation building and its relevance to disruption; the third section will present the methods employed for the laboratory experiment; the fourth part will present the results of the experiment; the final section will provide a discussion of the findings and the potential for generalisation and implications for policy.

With regard to the findings, it should be noted that this is a feasibility pilot study and the results should be viewed as suggestive, rather than conclusive. Further replication and expansion is required. The note's core value is in providing proof of concept for employing this methodology for cybercrime or related research, and offering insights for studies of these disruptive tactics, or others. We believe there is great value in this approach. Given social science experiments are very time-intensive and costly to run, it is important to report these pilot findings at this stage, as they may provide value to the research community and inform future studies using similar approaches.

Cybercriminal marketplaces

Cybercriminal marketplaces enable the sale of illegal goods and services online[9], including but not limited to: stolen or cloned credit cards, banking and personal information, narcotics, firearms, and cybercriminal services such as spam services, malware services, botnet services, and crimeware tutorials[10]. These platforms can also be used by cybercriminal networks as meeting places to identify/recruit potential co-offenders[11]. Many motivations are represented, and users access these platforms for different reasons and in different ways. For instance, a particular drug buyer may not be involved in broader criminality online, but simply seeks a ‘fix’ from any appropriate vendor. Meanwhile someone involved in the malware business may be engaged in a number of cybercriminal campaigns, and be very selective in terms of when they engage in a marketplace and with whom. While often presented as a new frontier, cybercrime marketplaces are surprisingly similar to existing legal online marketplaces such as eBay. They can show pictures of the products for sale, product details, as well as allowing for the possibility to contact vendors[12]. The particularity of criminal platforms, beyond their criminality, is that they often promise greater anonymity than legal platforms[13].

Recent attention has focussed on Darknet or cryptomarket sites, which are only accessible through the use of The Onion Router (Tor) and often make use of cryptocurrencies. Scholars have analysed perhaps the earliest cryptomarket Silk Road[14], the largest and most prolific to date AlphaBay[15], and the more recent leading cryptomarket DreamMarket[16]. These articles report that these sites cater mostly to drugs, that many transactions generate excellent feedback, that vendors are likely to use several aliases on one platform or trade on several platforms, and that drug offerings come from a few consumer countries rather than production countries[17].

But there also remain a range of Clearnet sites that span from the lowest to highest levels of criminality, which continue to play a very important role within the underground[18]. Identities are still guarded in these settings, through both social methods, such as the adoption of nicknames, and technical methods, such as the use of proxies. The more significant of these sites often have a financial (e.g. credit card data) or technical focus (e.g. malware), with drugs not being a major source of trade[19]. While the elite marketplace Darkode, a popular site for malware and other technical services, has been analysed in detail[20], many other cybercrime forums have not attracted quite the same scholarly attention as individual cryptomarkets.

A number of potential intervention strategies against these platforms have been discussed in the literature. These include attacking financial infrastructure or subverting physical deliveries[21]. But law enforcement responses to marketplaces, thus far, have largely been focussed on traditional tactics: arresting offenders and dismantling the sites. Such operations have been ongoing since the existence of early marketplaces like CarderPlanet, Shadowcrew and Darkmarket[22]. This approach continues to be in use, with four major platform takedowns in the first half of 2019 alone – xDedic[23], Wall Street Market and Valhalla/Silkkitie as part of Operation East River[24], and Deepdotweb[25].

While particular researchers have suggested that these operations may have had at least some effect in stunting the underground economy[26], several scholars have been sceptical of their long-term impact. For instance, the shutdown of Silk Road and other sites was followed by the creation of new platforms for online criminal trade[27]. The short-term effect of Operation Onymous – a large law enforcement action in November 2014 that targeted several cryptomarkets including Silk Road 2.0, Cloud 9, and Hydra – was the decrease in the total number of dealers registering on other platforms such as Agora and Evolution. But the number of active dealers recovered to almost pre-operation levels within one month, and two months after the operation the number of sales were double what they had been before. This suggests that the criminal trade was not eradicated but only displaced[28]. Ladegaard argues that media coverage of law enforcement operations, cybercriminal convictions and sentencing may even drive trade and vendor revenues to increase on other platforms[29]. While further research is required on this subject, it is clear that we need to develop a more thorough understanding of the mechanisms that would produce long-term harm to the criminal trade online.

Reputation and disruption

Given questions over existing tactics, not to mention the high level of resources they require, it seems sensible to examine other responses to cybercriminal marketplaces. One approach would be to better understand how the successful economic functioning of the markets could be damaged. In particular, one of their weaknesses appears to be the high level of uncertainty[30]. In their worst form, buyers are unable to ascertain the quality of the goods and services sold, and to verify the identity of market participants. Meanwhile vendors have no credible way of disclosing quality, are not regulated, and have been known not to provide the goods and services that have been paid for[31].

The inherent information asymmetry on these platforms creates what is known in economics as a 'market for lemons'. Akerlof famously notes that markets in which vendors cannot reliably signal trustworthiness and product quality may experience failure, as a greater number of vendors may have an incentive to offer products of low quality[32]. On the other hand, institutions like guarantees, brands, and licences can counteract such problems and allow trading to function more effectively. In line with existing findings, we contend that this is what happens in many underground marketplaces. In particular, rating systems have been put in place to counteract such challenges, which bear a strong resemblance to systems on eBay and other legal marketplaces[33]. Administrators of cybercriminal marketplaces have established other systems to mitigate identity and quality uncertainty and to convince buyers to trade with anonymous vendors[34]. Upon their arrival on particular platforms, vendors are encouraged to provide a sample of their product or service to administrators, or their appointees, in order for them to verify its genuine nature and quality. This initial vetting process then gives vendors the opportunity to display a 'reviewed vendor'[35], 'trusted vendor'[36], or 'verified status'[37] badge for all buyers to see.

Reputation is of vital importance to online cooperation in these markets[38]. Knowing the importance of reputation, these rating mechanisms could potentially be tampered with in order to encourage market failure. Two of the most discussed tactics for this purpose are slander and Sybil operations. Slander operations involve law enforcement and/or others creating fake profiles and leaving negative feedback on sellers in order to tarnish their reputation[39]. This should decrease the sellers' potential profits by increasing the perceived risk for buyers wanting to trade with them. Sybil operations consist of law enforcement and/or others creating fake profiles, which build up their reputation over time (perhaps by interacting with other fake profiles), but are used to default on sales with genuine buyers[40].

Recent experimental research shows that a lack of reputational information decreases cooperation in markets that rely on bilateral exchange[41]. By damaging the institutions that enhance cooperation, it is likely that these marketplaces revert back to the 'market for lemons' equilibrium. As a result, we would expect that, when these interventions are carried out:

(1) Prices will decrease, as buyers lack accurate information on product quality.

(2) Product quality will also decrease, as sellers have little incentive to market high quality products when low prices are predominant.

In the next section, we present the methods we used to test these expectations, and which we believe may be a broader benefit for the field as an approach for studying cybercrime disruption.

Methods

Although law enforcement agents have been able to infiltrate closed cybercriminal networks, it is much more complicated for researchers in academia to do so. Showing the necessary trustworthiness, technical abilities and 'credibility' to be accepted as part of such a network might involve taking part in criminal activities, including selling illicit goods or services[42]. While mere presence in particular groupings may create ethical quandaries, questions certainly would be raised around scholars directly carrying out interventions against cybercriminals, even in more open settings. Data on specific law enforcement interventions is also not widely available or gathered easily.

As a result, we have chosen to study this topic experimentally. Experiments have been consistently used in the social sciences for many decades, across economics, psychology and policy studies[43]. Like better-known scientific experiments, they aim to measure the effects of a 'treatment' (intentional change) on a group of human subjects compared to another group subjected to a control[44]. An advantage of this approach is that experimental designs can often be replicated and modified in future projects by the same or other researchers. Experiments are commonly carried out in a laboratory setting, though online experiments are growing in popularity. 'Field' experiments can be performed by making interventions in subjects' real lives, along with allowing for a control[45]. But as noted above, there are a number of ethical challenges in conducting a field experiment within cybercrime markets, along with limits to what interventions academic researchers (rather than law enforcement) could make.

In order to obtain a controlled setting to systematically test the impact of our interventions, we chose a novel application of a laboratory experiment. Within a laboratory setting, we replicated aspects of the online underground trade and disruptive interventions against it. While the experiment took place in a laboratory setting, the participants made use of computers to play the game. This experiment was coded in oTree, an increasingly popular platform in this research area (https://www.otree.org)[46]. One key benefit of oTree is that it is suited not only for laboratory experiments, but also for potential replication in an online experiment, which can scale to more participants and different populations. oTree is Python-based and provides a full end-to-end toolkit, handling the underlying logic and implementation of the games; the design, layout, and presentation of the user interface through which the participants interacted; the monitoring and processing of participant inputs; and finally, the capture, recording, and some basic processing of the raw experimental data. The applications used in our study consisted of modifications to the basic templates provided by those who developed the oTree platform and can be found at: http://otree-demo.herokuapp.com/demo/[47]. Our primary modification was the inclusion of a sending decision for vendors and rating system for buyers, as well as the disruption interventions that are central to the study.

The experimental design involved 144 participants. Given the laboratory we used only comfortably houses 24 people at one time, these participants were spread over 6 sessions. The last (control) session only managed to accrue 18 participants, leading to a total of 138 rather than 144 (accounted for in the below analysis). As is standard experimental practice, participants were paid for their involvement, with components of this payment also built into the game as incentives, which mimicked the profit-seeking behaviour of users of underground marketplaces[48]. Each session involved playing a trust game, a variation of a market for lemons game, and then filling out an end of experiment questionnaire to help the researchers better understand participant comprehension and engagement. While we provide a basic outline in the following paragraphs, further experiment details are provided in the appendices, through the participant information sheets. We also summarise the key details in Table 1.

For the trust game, each participant was first randomly paired up with another participant in the session. Participants remained anonymous to prevent reputation effects at this stage. The first mover decided what share of their 25 tokens to send to their counterpart; the counterpart then received triple that amount and decided how much to send back to the first mover. Participants were then randomly re-matched and played the other role, so they experienced both sides of the game, but without the possibility of punishing their pair partner from the previous round. This sequential game was similar to buyers paying for purchases without guarantee that the product will actually be sent and be of good quality, as it would happen in sites like eBay[49]. They were putting their blind trust in the other with no promise of return[50]. This game was used at the start of the session in order to measure trust between participants before the main game began, and was used again after it in order to evaluate how such trust evolved during the experiment.

Table 1. Experiment Summary.

Element	Number
Total Participants	138 (up to 24 in each session)
Sessions	6 (2 sessions for each treatment)
Rounds	30 rounds in each session

Table 2. Product grades, costs to vendors, and values to buyers.

Grade	High	Medium	Low
Production cost	30	20	10
Value to buyer	45	30	15

Buyer payoffs: 50 tokens + value of the grade purchased (if sent) – vendor's price

Vendor payoffs: 50 tokens + vendor's price – cost of the grade produced (if sent)

Table 3. Percentage of buying and sending decisions throughout rounds and across all Treatments.

Percentage of buying and sending decisions throughout rounds and across all Treatments			
	Sent	Not sent	Not bought
CONTROL	91	8	1
SLANDER	90	9	1
SYBIL	92.5	7	0.5

The ‘market for lemons’ component, which was the heart of this experiment, was based on Holt and Sherman’s classroom game[51] and the more recent oTree template[52]. Participants were randomly assigned a fixed role throughout this task – buyer or vendor. For each round, groups of three were randomly allocated, containing one buyer and two vendors. They played the round and then were randomly re-matched into new groups to continue in the next round. This re-matching reduced any potential for vendor collusion as part of a duopoly, which has been criticised in the experimental literature[53].

At the beginning of each trading round, all the participants received 50 tokens. Vendors began by privately choosing a price and a quality grade for their products. The grade could be high, medium, or low, with higher grades costing more to produce and worth more to buyers. Buyers then had a chance to purchase from one of the vendors at the price listed or not to buy anything. Before purchase, buyers could only observe price, and could not determine quality grades. After purchase, the grade of the bought unit was revealed to them. They would then provide a rating on the seller, which would be visible to future potential buyers and updated after each round. While the instructions provided clear details, and parallels with an online marketplace may have been obvious, for this pilot no specific reference was made to this setting replicating cybercrime – thereby avoiding some ethical pitfalls (such as any trepidation that we are promoting criminal activity or ‘training’ participants in it) and increasing external validity beyond this particular cybercriminal setting.

Payoffs for vendors and buyers are summarised in Table 2. Participants only had access to their own role’s payoff calculations.

Several iterations of this trading game were played by participants in 6 different laboratory sessions, grouped into three treatments, with each involving 30 rounds:

a) CONTROL: Two control sessions involved no interventions: vendors advertised prices, buyers chose whether to buy from a vendor or not, vendors selected whether to send the product or default, the buyers rated the purchases they had just made after the quality grade of the product was revealed (see Figure 1).

b) SLANDER (Treatment 1): These two sessions simulated the slander attack, by introducing a 15% chance that each buyer’s rating was ‘compromised’, and replaced with a randomly selected rating, during each round. Participants were informed of this rate of compromise at the start of the experiment but were not informed when they had been affected by it either as a buyer or vendor (see Figure 2).

c) SYBIL (Treatment 2): The remaining two sessions simulated the Sybil attack, by introducing a 15% chance that each vendor’s positive ‘sending’ decision was randomly ‘compromised’, and not in fact sent, during each round. Participants were informed of this rate of compromise at the start of the experiment but were not informed when they had been affected by it either as a buyer or vendor (see Figure 3).

Results

In this section, we provide results on the impact of the slander and Sybil interventions on prices, quality grades and buying/sending decisions. Here we make use of descriptive, rather than inferential, statistics. The reason for this is that our primary aim is to

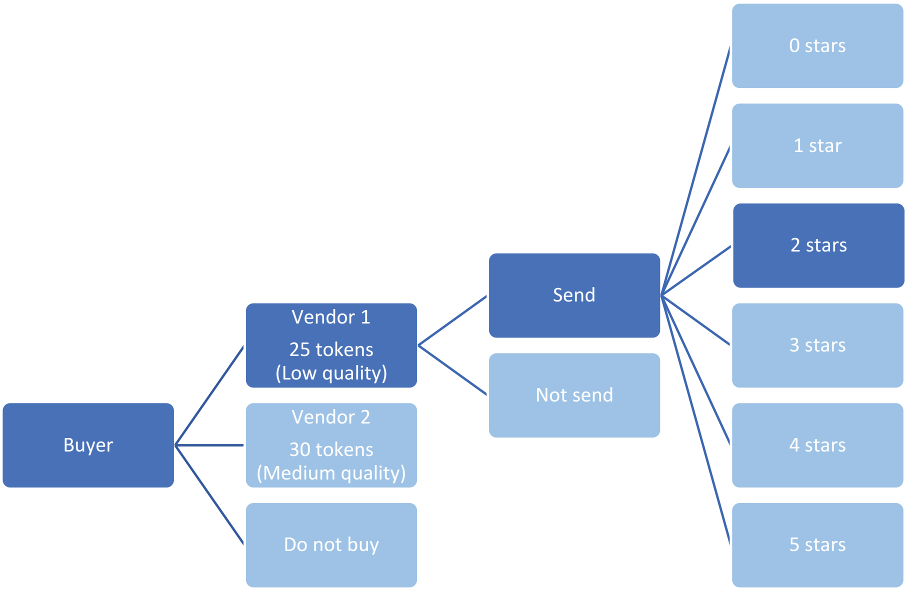

Figure 1. Example of a Control decision sequence.

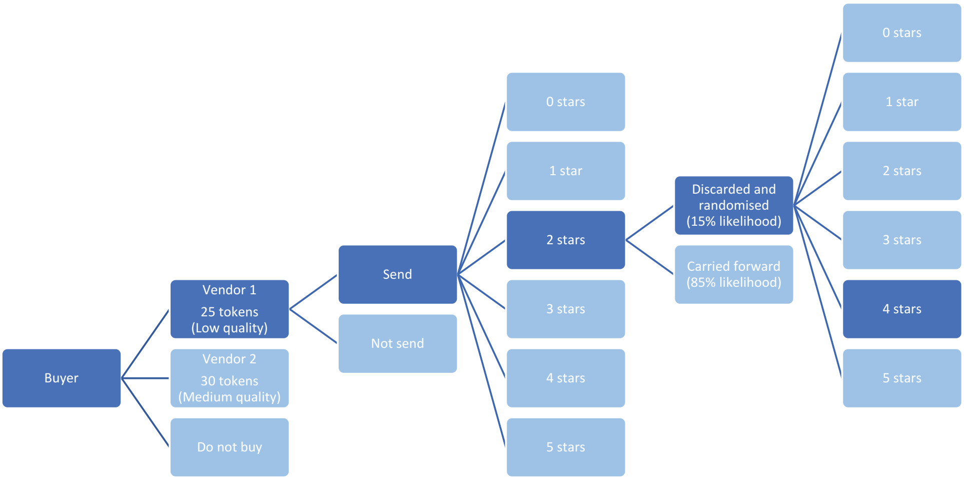

Figure 2. Example of a Slander decision sequence.

demonstrate feasibility and to broadly illustrate the nature of the experimental process and its outcomes. Simple significance tests cannot be used in this case, as the data points are not independent because subjects interacted through group allocations. While more complex quantitative analysis is possible on this form of data, this is better suited to a larger sample size with greater statistical power than what we gathered for this pilot. Such a plan would analyse the full experimental data with non-parametric statistics to examine treatment effects at the session level, and regression analysis to investigate individual behaviour.

Prices offered

The average prices offered by vendors in CONTROL (the Control) was 21.3 tokens, compared to 16.8 tokens in SLANDER (Treatment 1) and 19.9 tokens in SYBIL (Treatment 2). This was a price drop of 4.5 tokens in relation to the slander attack, and 1.4 for the Sybil attack. Prices were driven downwards throughout the experiment, with stronger effects in SLANDER and then SYBIL. Importantly, both the Treatments and the Control started with similar prices, which indicates a dynamic process of learning and operating in a compromised market, rather than initial differences between the groups (See Figure 4.).

Across all Treatments, one can observe there were two common pricing strategies. Some priced Low (as low as 0 or 5 tokens), likely to encourage buyers to make a purchase despite uncertain qualities. The need for vendors to make a sale to have a chance of earning payoffs (compared to getting 0 tokens if they were not chosen for the sale) may have encouraged some to sell products at a small margin in order to create an initial reputation for themselves and help secure future purchases based on their rating. Many of these low-priced products were produced at Medium or High qualities. The other strategy was to price High (as high as 45 or 50 tokens), possibly in an attempt to signal quality to buyers.

Purchase prices

Although the prices at which buyers decided to buy were dependent on the prices offered by vendors, buyers still had a choice between two offers during each round. It is therefore of some interest to understand which price points buyers chose for their purchases. The average price chosen by buyers in CONTROL was 20.5 tokens, compared to 14.5 tokens in SLANDER and 18.4 tokens in SYBIL, representing a price drop of 6 tokens and 2.1 tokens respectively. These are below the average prices offered by vendors in each treatment, suggesting that buyers often chose the cheaper offers. In line with the above discussion,

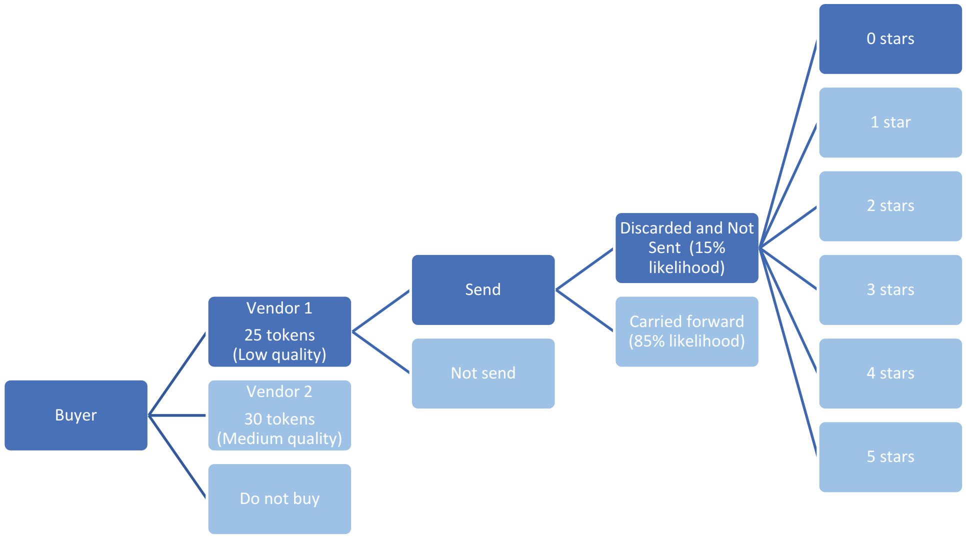

Figure 3. Example of a Sybil decision sequence.

the prices at which buyers bought products also decreased throughout rounds. Between the first and last rounds, prices chosen decreased by 42% in CONTROL, 39.6% in SLANDER, and 46.9% in SYBIL (See Figure 5.).

Quality

There was a decrease in quality when the interventions were applied, but the effect was not as pronounced as it was for price. SLANDER saw Low quality products being sold most commonly, compared to the predominance of Medium quality products when the Sybil intervention was applied. CONTROL were the only sessions in which High quality grades were produced in higher quantities, on average, than Medium and Low ones, in that order. SLANDER sessions showed the opposite situation in which Low quality grades were the ones most produced, ahead of Medium and High ones, in that order. Finally, Medium quality grades were the most produced in SYBIL sessions, with High and Low ones reaching similar averages (See Figure 6.). There did not appear to be a constant decrease in quality production, rather quality production stagnated at these levels for all three grades throughout rounds.

Buying and sending decisions

Similar results were seen across Treatments in terms of buying and sending decisions, which are summarised in the Table 3

.Very few buyers chose not to buy from either vendor, due to them being able to make more money from buying even a low-quality product than forfeiting their tokens at the end of the round. There were 4 rounds in the Control sessions when 1 buyer chose not to buy, 6 rounds in SLANDER, and 2 rounds in SYBIL. These rounds were most often situated at the beginning of sessions and are therefore not necessarily seen to reflect a loss of trust between participants, but rather participants testing the system or making sense of the game..

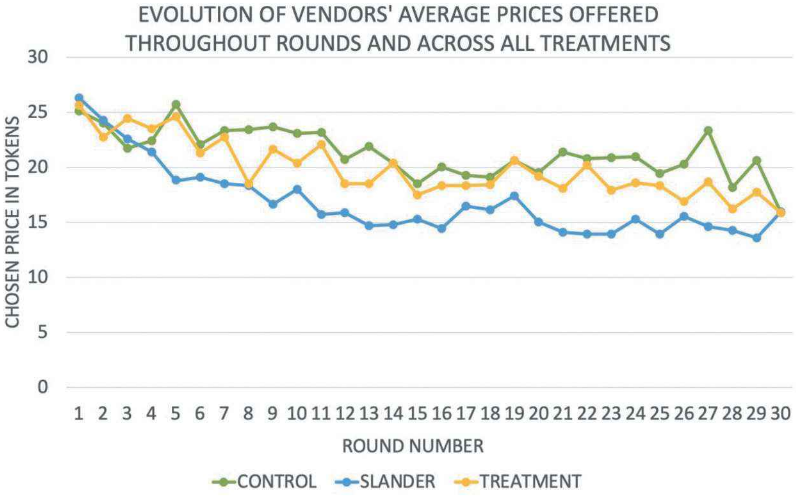

Figure 4. 'Market for lemons' evolution of average prices offered by vendors throughout rounds and across treatments.

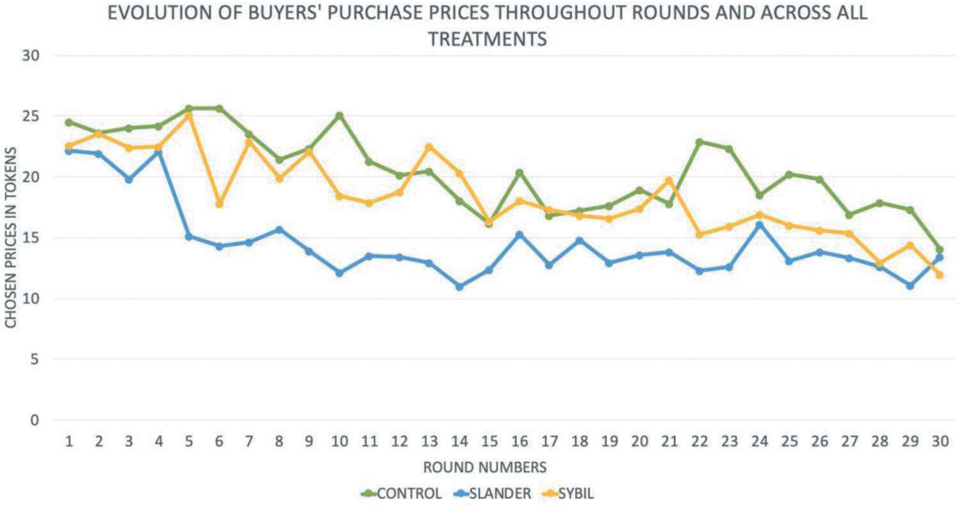

Figure 5. 'Market for lemons' evolution of average purchase prices throughout rounds and across treatments.

Overall, vendors seemed to consistently send their products, even in SYBIL, where this intervention might have provided cover for further defaulting. Only the last two rounds saw an increase in decisions not to send, which is the typical endgame effect as vendors try to maximise their payoffs when they no longer fear reputational consequences (See Figure 7.).

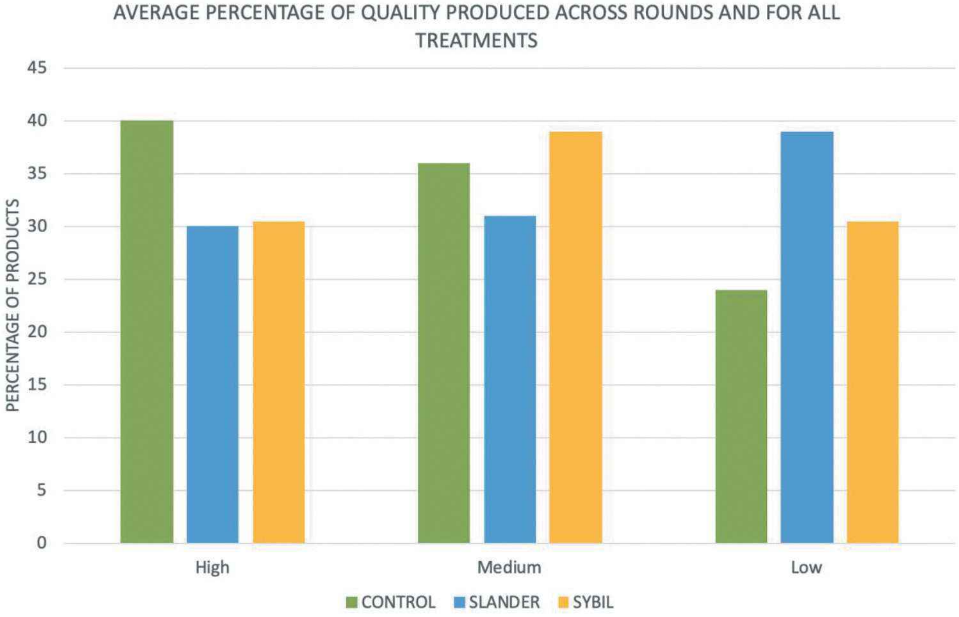

Figure 6. 'Market for lemons' comparison of average quality offered by vendors across treatments.

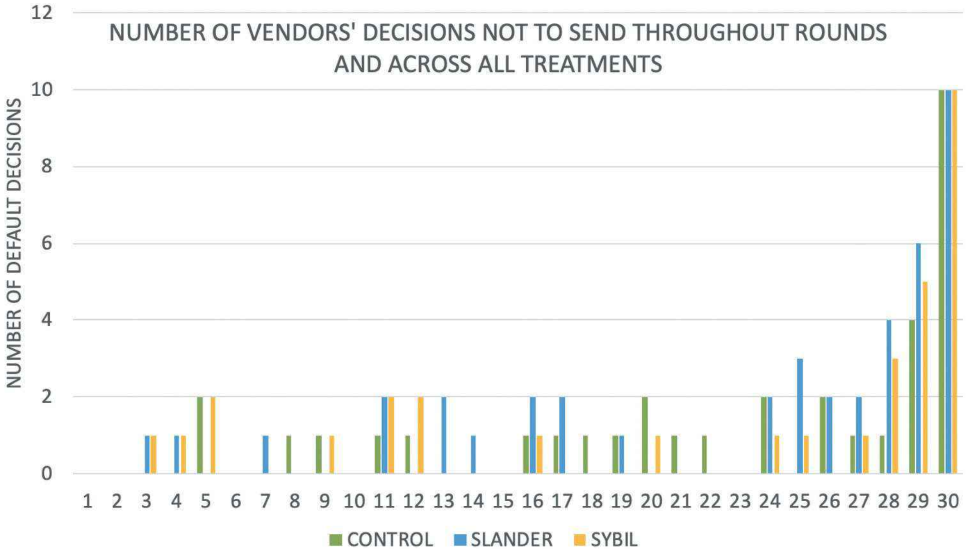

Figure 7. 'Market for lemons' evolution of vendors' decisions not to send throughout rounds and across treatments.

Trust

In this section we examine whether the interventions had any impact on broader measures of trust. All the participants played the trust game before the market for lemons game, and then again afterwards. In the first instance of the trust game, similar averages could be seen across sessions in the number of tokens sent and sent back between participants. Tokens sent ranged from 0 to 25 with averages between 13 and 15 across sessions. Tokens sent back varied across a wider range, often from 0 to 50 tokens, though they were generally below the number of tokens initially sent, averaging between 12 and 14, and were more likely to include 0 tokens than the initial sending decisions. These findings were comparable across Treatments.

In the final trust game, following the market for lemons, similar averages could still be seen. Tokens sent ranged from 0 to 25, and averaged between 11.5 and 12.5. Overall this was therefore a small decrease from the first trust game: 9% in CONTROL and SYBIL; and 17.5% in SLANDER. Tokens sent back again varied across a wider range, and in only one instance were actually above the number of tokens initially sent. In this case, averages lay between 9 and 14, which was again an overall decrease from the first game across Treatments. There was a 25% and 22% decrease respectively in CONTROL and SLANDER in the amounts sent back. There was a slight increase of 4% in the amounts sent back in SYBIL.

Discussion

The pilot findings presented in the previous section paint a picture of the interventions leading to a low-price and low-quality market. This suggests preliminary confirmation of both of our expectations. As such, it would support the broader theory advanced by Akerlof, particularly in relation to the importance of institutions[54]. In this case, by targeting the reputation systems that enhance cooperation, we damaged the trade. It would appear that with a less accurate way for buyers to determine quality, prices drop, and therefore so will quality. The slander attack seemed to have a greater impact. This is interesting in and of itself, but also because this intervention directly targets the reputation mechanism, whereas the Sybil attack does this more indirectly through increasing defaults which may then lead to lower ratings.

In evaluating these results, validity is one of the most important topics to discuss. Laboratory experiments bring great advantages in terms of applying the scientific method to social and economic phenomena. But questions remain as to how well such experiments are matched to reality. Our initial experimental design is intentionally broad to allow for the development of a method, and findings, that may be widely generalisable to extra-legal markets. But as a pilot and first step, we accept that we cannot incorporate all the rich detail of a particular cybercriminal marketplace or its user-base, without also introducing considerable noise and so many variables that it would become impossible to determine the relative importance of each in relation to cooperation and its disruption. Robust experiments are simple by design. In our case we have focussed on replicating the most fundamental element of cybercriminal marketplaces through the market for lemons game, and the selection of two of the most widely discussed interventions – the Sybil and slander attacks.

While some detailed aspects of cybercriminal marketplaces cannot be included in this pilot experiment, these components could be gradually incorporated as part of a significant work program involving multiple sequential experiments. By adding significant features one by one, the impact of each variable can be accounted for systematically, with the potential for noise drastically reduced[55]. Future iterations could be matched more closely to the nature of specific users and marketplaces. For instance, participants could be informed about the cybercriminal nature of the platform and assigned roles based on the specific motivations of users within the underground economy. Another design extension might allow for the trade of named products and for the game to take into account the different dynamics of, for example, technical products of a virtual nature as opposed to physical drug packages, and the stakes involved in each case. The known behavioural patterns of peers within particular offender environments, for instance around rating and reputation, might also be built into different variations of the experiment. Significant institutional elements, such as escrow providers who are employed for certain cybercriminal transactions, might be included[56].

A particularly important design element to consider in future experiments is the sample of participants. Following standard convention within social science laboratory experiments, our pilot made use of a sample of university students[57]. A sensible critic might question how well this sample matches the cybercriminal population and suggest this is a major limitation. Our response is in three parts. First, particularly within this early pilot phase, we seek a more foundational design that speaks to extra-legal markets as a broader category, rather than cybercrime alone. While there is enough promise in this pilot to spark more focussed cybercrime experiments, the most sensible initial design is to hew closely to convention before innovating with the sample. Second, there are standard justifications in the experimental field as to the wider question of student samples and validity. A key one is not to assume that student samples will necessarily differ in their behaviour from other samples or populations. While there are certainly well-known examples of this, the best general practice is to employ a second sample of non-students for comparison.[58] Therefore, as part of follow up experimental designs, a sensible next step would be to employ a non-student sample, with a possible further step of selecting demographic characteristics, which are specifically matched to cybercriminal profiles. Finally, ethics applications and logistics allowing, it might even be possible to play such games with a sample of former offenders. But this a last step not a first step. Former cybercriminals are a small and hard to reach population; they would require significant time/resources to recruit. The population is also not monolithic, so some sampling issues would remain, or a further sequence of experiments would need to account for this variation[59]. Given the complexities involved, this final step is best taken once an experimental design has been perfected over a number of iterations. While it remains an exciting and ambitious aim on the horizon, it is well beyond the scope of this pilot paper.

Another aspect of validity to discuss is how well matched the experimental design is to real world police behaviour and successful disruption strategies. In this experiment, we chose a compromise rate of 15%, as this seemed substantial without being hugely unrealistic in operational terms. It is also in line with existing cybercriminal behaviour. For instance, Espinosa estimates that between 83% and 88% of vendors actually 'send' their orders on platforms such as Hansa[60]. Adding in another 15% of defaults would therefore double the expected default rate within such a marketplace. But it may be possible that this level is not feasible for undercover agents to facilitate in real cybercriminal marketplaces of large size. Lower compromise rates could be tested and/or tactics could be explored concerning how to make the interventions better targeted to particular individuals who play an outsize role in trading, or smaller elite markets. Another avenue would be to investigate if these interventions can be automated in some way.

There is also the question of evaluating the potential success of these disruptive interventions in the wild. A reasonable hypothesis would be that driving prices and quality down in an online illicit market should limit the profits of cybercriminals and discourage their participation, at least in the marketplace, if not in their broader criminal activity. But such a hypothesis requires further testing. Whether a low-price and low-quality market is a good measure of success in fighting cybercrime, might depend on a variety of factors. For one, relatively high levels of harm might continue as such a market might still have a high volume of trade. The low prices might encourage this in some ways. It is also likely that the type of products traded in a given marketplace will relate to the level of harm. A marketplace that is forced to sell low quality malware or financial data, even at a high volume, may be causing limited harm. But a marketplace that sells low quality drugs, may cause significant harm, perhaps greater than a market that functions optimally and in so doing trades in much safer products. Finally, questions remain around whether such interventions would have a long-term impact, or might only frustrate cybercriminals in the short-term, or require continued operations in order for the level of disruption to be maintained[61].

Results from this experiment and related efforts could have important policy contributions. These methods can be used to investigate topics that are very hard for researchers to study in the 'wild' and to study them in a controlled way, which is bound by the scientific method. In this case, it has allowed for the study of disruptive interventions that have not been the subject of wide reporting or empirical testing. It has also allowed for an assessment of which of the two interventions might be more effective. In this case, it appears to be the slander attack, which conveniently happens to be cheaper and easier to carry out than the Sybil attack, as well. Ultimately, such findings need to be tested as part of larger experiments, but also operationally by police and others to know how effective they could be and to ensure that practical tactics for their deployment are optimised. But the adaptation of experimental methods to cybercrime offers an exciting research direction going forward.

Conclusion

Inspired by cybercrime, this research note presented the pilot results of a social laboratory experiment modelling an extra-legal marketplace, and interventions against it. The two simulated interventions were the slander attack and the Sybil attack. While the slander attack appeared more impactful, both tactics showed some effectiveness in making the marketplace one of low price and low quality. By targeting the reputation mechanism, these attacks appeared to revert the setting back to a ‘market for lemons’ equilibrium. This suggests these tactics may be fruitful in fighting cybercrime (or crime more generally), by making it less efficient, which in turn should reduce profits.

Given this was a feasibility pilot, the results should be treated as suggestive. There is much room for carrying out further experiments with larger sample sizes and greater statistical power. There is also room for building in more specific cybercrime details tied to particular real marketplaces and their users, or expanding the scope of this experiment to incorporate components such as trade volume and/or whether there is a tipping point at which cybercriminals could be nudged out of the business altogether. There is scope for these approaches to be tested by law enforcement and/or industry in real cybercrime markets, as part of a field experiments that could lead to other interesting findings.

This research note’s primary contribution is the illustration of a method that has not been applied to the study of cybercrime markets. While not without its own challenges, adopting experimental approaches might help overcome a number of limitations in the field, and particularly with regard to the study of disruption. It offers a way to examine this topic empirically, when other data is not available. It also offers a strong application of the scientific method, which means that statements can potentially be made around causation, allowing a clearer assessment of the impact of particular interventions. This could be of enormous value to policy discussions. Nonetheless, this experimental approach is intricate, time-consuming and expensive. Thus considerable thought and planning is required to produce future experimental studies that are tightly matched to studying cybercrime disruption. We hope that this research note provides both some encouragement and some learnings for those researchers who may take up this challenge in the future with regard to these interventions, or others.

Notes

1. Décary-Hétu and Giommoni ‘Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Onymous’.

2. Décary-Hétu and Giommoni ‘Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Onymous’; Van Buskirk et al, ‘The Closure of the Silk Road: What Has This Meant for Online Drug Trading?’; Barratt et al, ‘Safer Scoring? Cryptomarkets, Social Supply and Drug Market Violence’.
3. EC3, ‘Internet Organised Crime Threat Assessment’.
4. Van Buskirk et al, “The Closure of the Silk Road: What Has This Meant for Online Drug Trading?; Europol, ‘Deepdotweb Shut Down: Administrators Suspected of Receiving Millions of Kickbacks from Illegal Dark Web Proceeds’.
5. Décary-Hétu and Giommoni ‘Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Onymous’; Van Buskirk et al, ‘The Closure of the Silk Road: What Has This Meant for Online Drug Trading?’; Yip et al, ‘Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing’; Hutchings et al, ‘Taking Down Websites to Prevent Crime’; Ladegaard, ‘We Know Where You Are, What You Are Doing and We Will Catch You: Testing Deterrence Theory in Digital Drug Markets’; Paquet-Clouston, ‘Cybercrime Is Whose Responsibility? A Case Study of an Online Behaviour System in Crime’.
6. Yip et al, ‘Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing’; Décary-Hétu and Dupont, ‘Reputation in a Dark Network of Online Criminals’; Hutchings and Holt. ‘The Online Stolen Data Market: Disruption and Intervention Approaches’; Paquet-Clouston et al, ‘Assessing Market Competition and Vendors’ Size and Scope on Alphabay’.
7. Kagel and Roth, ‘*The Handbook of Experimental Economics’*.
8. Maimon et al ‘Restrictive Deterrent Effects of a Warning Banner in an Attacked Computer System’.
9. Décary-Hétu and Giommoni ‘Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Onymous’; Décary-Hétu and Dupont, ‘Reputation in a Dark Network of Online Criminals’.
10. Kigerl, ‘Profiling Cybercriminals: Topic Model Clustering of Carding Forum Member Comment Histories’.
11. Leukfeldt et al, ‘Origin, Growth and Criminal Capabilities of Cybercriminal Networks. An International Empirical Analysis’; Huang et al, ‘Casting the Dark Web in a New Light’.
12. Bartlett, ‘*The Dark Net’*.
13. Lusthaus, ‘Trust in the World of Cybercrime’.
14. Christin, ‘Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace’; Barratt et al ‘Use of Silk Road, the Online Drug Marketplace, in the United Kingdom, Australia and the United States’.
15. Paquet-Clouston, Masarah, David Decary-Hetu, and Carlo Morselli. ‘Assessing Market Competition and Vendors’ Size and Scope on Alphabay’.
16. Zhou et al, ‘A Market in Dream: The Rapid Development of Anonymous Cybercrime’.
17. Christin, ‘Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace’; Soska et al, ‘Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem’.
18. Hutchings et al, ‘Taking Down Websites to Prevent Crime’; Lusthaus, ‘*Industry of Anonymity: Inside the Business of Cybercrime’*.
19. Yip et al, ‘Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing’; Lusthaus, ‘Trust in the World of Cybercrime’.
20. Dupont et al, ‘Darkode: Recruitment Patterns and Transactional Features of “the Most Dangerous Cybercrime Forum in the World”’.
21. Christin, ‘Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace’.
22. Lusthaus, ‘Industry of Anonymity: Inside the Business of Cybercrime’; Poulsen, ‘Kingpin’.
23. Europol, ‘Xdedic Marketplace Shut Down in International Operation’.
24. Europol, ‘Double Blow to Dark Web Marketplaces’.
25. Europol, ‘Deepdotweb Shut Down’.

26. Bradley and Stringhini, 'A Qualitative Evaluation of Two Different Law Enforcement Approaches on Dark Net Markets'; Bhaskar et al, 'The Economic Functioning of Online Drugs Markets'.
27. Décary-Hétu and Giommoni 'Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Onymous'; Van Buskirk et al, 'The Closure of the Silk Road: What Has This Meant for Online Drug Trading?'; Yip et al, 'Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing'; Hutchings et al, 'Taking Down Websites to Prevent Crime'; Ladegaard, 'We Know Where You Are, What You Are Doing and We Will Catch You: Testing Deterrence Theory in Digital Drug Markets'; Paquet-Clouston, 'Cybercrime Is Whose Responsibility? A Case Study of an Online Behaviour System in Crime'.
28. Décary-Hétu and Giommoni 'Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Onymous'; Van Buskirk et al, 'The Closure of the Silk Road: What Has This Meant for Online Drug Trading?'; Bradley and Stringhini, 'A Qualitative Evaluation of Two Different Law Enforcement Approaches on Dark Net Markets'; Bhaskar et al, 'The Economic Functioning of Online Drugs Markets'.
29. Ladegaard, 'We Know Where You Are, What You Are Doing and We Will Catch You: Testing Deterrence Theory in Digital Drug Markets'.
30. Yip et al, 'Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing'.
31. Perhaps the first to note the connection between cybercriminal marketplaces and markets for lemons were Herley and Florêncio, 'Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy'. On the broader theoretical point, see Dasgupta, 'Trust as a Commodity'.
32. Akerlof, 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism'. Many of the discussions in this paper relate to the concept of trust, as well as cooperation (which can take place in the absence of trust). It is beyond the scope of this paper to engage in an extensive review of the theoretical literature, but we follow James Coleman's approach to trust, whereby trust is: 'an incorporation of risk into the decision of whether or not to engage in the action. This incorporation of risk into the decision can be treated under a general heading that can be described by the single word "trust." Situations involving trust constitute a subclass of those involving risk. They are situations in which the risk one takes depends on the performance of another actor' (Coleman, 'Foundations of Social Theory', p. 91). See also Dasgupta, 'Trust as a Commodity'.
33. Christin, 'Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace'; Bhaskar et al, 'The Economic Functioning of Online Drugs Markets'; Resnick and Zeckhauser. 'Trust among Strangers in Internet Transactions: Empirical Analysis of Ebay's Reputation System'; Dellarocas, 'The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms'; Melnik and Alm, 'Does a Seller's Ecommerce Reputation Matter?'; Li, 'Reputation, Trust, and Rebates: How Online Auction Markets Can Improve Their Feedback Mechanisms'; Jin and Kato, 'Price, Quality, and Reputation: Evidence from an Online Field Experiment'; Houser and Wooders, 'Reputation in Auctions: Theory, and Evidence from Ebay'.
34. Ibid note 30.
35. Ibid note 30.
36. Décary-Hétu and Laferrière, 'Discrediting Vendors in Online Criminal Markets'.
37. Dupont et al, 'The Ecology of Trust among Hackers'.
38. Lusthaus, 'Industry of Anonymity: Inside the Business of Cybercrime'.
39. Mell, '*Promoting Market Failure: Fighting Crime with Asymmetric Information'*.
40. Yip et al, 'Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing'; Paquet-Clouston, 'Cybercrime Is Whose Responsibility? A Case Study of an Online Behaviour System in Crime'; Décary-Hétu and Dupont, 'Reputation in a Dark Network of

Online Criminals'; Hutchings and Holt, 'The Online Stolen Data Market: Disruption and Intervention Approaches'; Décary-Hétu and Laferrière, 'Discrediting Vendors in Online Criminal Markets'.

41. Gallo and Yan, 'The Effects of Reputational and Social Knowledge on Cooperation'.
42. Leukfeldt et al, 'Origin, Growth and Criminal Capabilities of Cybercriminal Networks. An International Empirical Analysis'; Lusthaus, 'Trust in the World of Cybercrime'.
43. Rivlin, 'Social Experiments: Their Uses and Limitations'.
44. Heckman and Smith, 'Assessing the Case for Social Experiments'.
45. Greenberg and Shroder, '*The Digest of Social Experiments'.*
46. oTree website, 'oTree Demonstrations – Lemon Market Game Session'.
47. Ibid note 45.
48. Webster and Sell, '*Laboratory Experiments in the Social Sciences*'; Von Neumann and Morgenstern '*Theory of Games and Economic Behavior'.*
49. Resnick and Zeckhauser, 'Trust among Strangers in Internet Transactions: Empirical Analysis of Ebay's Reputation System'.
50. Berg et al, 'Trust, Reciprocity and Social History'.
51. Holt and Sherman 'Classroom Games: A Market for Lemons'.
52. Ibid note 45.
53. Huck et al, 'Two Are Few and Four Are Many: Number Effects in Experimental Oligopolies'.
54. Ibid note 32.
55. With its own particular strengths and limitations, agent-based simulation is another approach capable of approximating interventions against cybercriminal settings, and which could be used to build in some real-world elements, such as scale, which would be far too expensive to model in laboratory or online experiments. See, for example, Duxbury and Haynie 'Criminal Network Security: An Agent-Based Approach to Evaluating Network Resilience'; Duxbury and Haynie, 'The Responsiveness of Criminal Networks to Intentional Attacks: Disrupting Darknet Drug Trade'.
56. The empirical literature on cybercrime provides a strong foundation for determining many of these details. See, for instance, Yip et al, 'Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing'; Lusthaus, '*Industry of Anonymity: Inside the Business of Cybercrime*'; Christin, 'Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace'; Leukfeldt et al, 'Origin, Growth and Criminal Capabilities of Cybercriminal Networks. An International Empirical Analysis'.
57. See, for instance, the student samples across a range of experiments in Camerer, '*Behavioral Game Theory: Experiments In Strategic Interaction'.*
58. On student sampling see Druckman and Kam 'Students as Experimental Participants: A Defense of the "Narrow Data Base"'. For a famous example of diverse game behaviour across particular samples, see Henrich et al, 'In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies'.
59. On the range of cybercriminal profiles, see footnote 38, 10-17.
60. Espinosa, 'Scamming and the Reputation of Drug Dealers on Darknet Markets'.
61. Short-lived success has been a criticism of takedown and arrest operations against online marketplaces. See footnote 5.

Acknowledgement

A preliminary version of this paper was presented at the 3rd Workshop on Attackers and Cyber-Crime Operations in 2021. Ivan Flechais and Andrew Mell provided some valuable insights during the conceptualisation of this project, while Faisal Hameed and Thomas Burton completed student scoping projects as part of this initial phase. We thank Raymond Duch and Noah Bacine for hosting the experiment at the Centre for Experimental Social Sciences at Nuffield College. It could not have been completed without their support and expertise. In terms of co-author responsibilities: Lonie Sebagh was the lead author, and primarily responsible for designing and carrying out the pilot

experiment. Jonathan Lusthaus was the Principal Investigator, coordinating and involved in all aspects of the study. Edoardo Gallo and Federico Varese were centrally involved in conceptualisation, research design and editing the paper. Sean Sirur provided technical assistance in programming the experiment.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a grant from the Cisco University Research Program Fund, a corporate advised fund of Silicon Valley Community Foundation (CG ID#: 1376716).

Notes on contributors

Lonie Sebagh is a doctoral candidate in Cyber Security in the Department of Sociology, University of Oxford. Her research is on the disruption of the online criminal trade by various stakeholders from law enforcement to private industry, non-profits, government, and trading platforms combining sociology, criminology, and economics methods and perspectives.

Jonathan Lusthaus is Director of The Human Cybercriminal Project and a Senior Research Fellow in the Department of Sociology, University of Oxford. He is also a Research Fellow at Nuffield College. His research focuses on the ″human″ side of profit-driven cybercrime: who cybercriminals are and how they are organised.

Edoardo Gallo is a University Associate Professor at the Faculty of Economics and an Official Fellow in Economics at Magdalene College, University of Cambridge. His research interests include Experimental Economics, Behavioural Economics and Networks.

Federico Varese is Head of the Department of Sociology, Professor of Criminology and a Senior Research Fellow at Nuffield College, University of Oxford. His main area of research is the study of organised crime.

Sean Sirur is a doctoral candidate in Cyber Security in the Department of Computer Science, University of Oxford. His main research interests are in trust and reputation.

Bibliography

Akerlof, G. “The Market for “Lemons”: Quality Uncertainty and the Market Mechanism.” *The Quarterly Journal of Economics* 84, no. 3 (1970): 488–500.

Barratt, M., J. Ferris, and A. Winstock. “Use of Silk Road, the Online Drug Marketplace, in the United Kingdom, Australia and the United States.” *Addiction* 109, no. 5 (2014): 774–783.

Barratt, M., J. Ferris, and A. Winstock. “Safer Scoring? Cryptomarkets, Social Supply and Drug Market Violence.” *International Journal of Drug Policy* 35 (2016): 24–31.

Bartlett, J. *The Dark Net*. London: William Heinemann, 2014.

Berg, J., J. Dickhaut, and K. McCabe. “Trust, Reciprocity and Social History.” *Games and Economic Behavior* 10, no. 1 (1995): 122–142.

Bhaskar, V., R. Linacre, and S. Machin. “The Economic Functioning of Online Drugs Markets.” *Journal of Economic Behavior & Organization* 159 (2019): 426–441.

Bradley, C., and G. Stringhini. “A Qualitative Evaluation of Two Different Law Enforcement Approaches on Dark Net Markets.” In *WACCO*. Stockholm: IEEE, 2019.

Chen, D. L. “oTree Demonstrations - Lemon Market Game Session.” Accessed March 1st, 2019. http://otree-demo.herokuapp.com/p/alq45z1c/lemon_market/Introduction/1/

Christin, N. “Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace.” In *The World Wide Web Conference*. Rio de Janeiro: ACM, 2013.

Dasgupta, P. “Trust as a Commodity.” In *Trust: Making and Breaking Cooperative Relations*, edited by D. Gambetta, 49–72. Oxford: Basil Blackwell, 1988.

Décary-Hétu, D., and B. Dupont. “Reputation in a Dark Network of Online Criminals.” *Global Crime* 14, no. 2–3 (2013): 175–196.

Décary-Hétu, D., and L. Giommoni. “Do Police Crackdowns Disrupt Drug Cryptomarkets? A Longitudinal Analysis of the Effects of Operation Onymous.” *Crime, Law and Social Change* 67, no. 1 (2017): 55–75.

Décary-Hétu, D., and D. Laferrière. “Discrediting Vendors In Online Criminal Markets.” In *Disrupting Criminal Networks: Network Analysis in Crime Prevention*, edited by A. Malm and G. Bichler, 129–152. Boulder: Lynne Rienner, 2015.

Dellarocas, C. “The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms.” *Management Science* 49, no. 10 (2003): 1407–1424.

Dittus, M., J. Wright, and M. Graham. “Platform Criminalism: The ‘Last-mile’ Geography of the Darknet Market Supply Chain.” In *The World Wide Web Conference*. Lyon: ACM, 2018.

Dupont, B., A.-M. Côté, J.-I. Boutin, and J. Fernandez. “Darkode: Recruitment Patterns and Transactional Features of “The Most Dangerous Cybercrime Forum in the World”.” *American Behavioral Scientist* 61, no. 11 (2017): 1219–1243.

Dupont, B., A.-M. Côté, C. Savine, and D. Décary-Hétu. “The Ecology of Trust among Hackers.” *Global Crime* 17, no. 2 (2016): 129–151.

Espinosa, R. “Scamming and the Reputation of Drug Dealers on Darknet Markets.” *International Journal of Industrial Organization* 67 (2019): 1–26.

Europol. “Deepdotweb Shut Down,” Europol. Accessed July 16, 2020a. https://www.europol.europa.eu/newsroom/news/deepdotweb-shut-down-administrators-suspected-of-receiving-millions-of-kickbacks-illegal-dark-web-proceeds .

Europol. “Deepdotweb Shut Down: Administrators Suspected of Receiving Millions of Kickbacks from Illegal Dark Web Proceeds,” Europol. Accessed July 16, 2020b. https://www.europol.europa.eu/newsroom/news/deepdotweb-shut-down-administrators-suspected-of-receiving-millions-of-kickbacks-illegal-dark-web-proceeds .

Europol. “Double Blow to Dark Web Marketplaces,” Europol. Accessed July 16, 2020c. https://www.europol.europa.eu/newsroom/news/double-blow-to-dark-web-marketplaces .

Europol. “xDedic Marketplace Shut down in International Operation.” Europol. Accessed July 16, 2020d. https://www.europol.europa.eu/newsroom/news/xdedic-marketplace-shut-down-in-international-operation .

Europol. *Internet Organised Crime Threat Assessment*. The Hague, 2019.

Gallo, E., and C. Yan. “The Effects of Reputational and Social Knowledge on Cooperation.” *PNAS* 112, no. 12 (2015): 3647–3652.

Greenberg, D. H., and M. Shroder. *The Digest of Social Experiments*. Third ed. Washington, D.C: Urban Institute Press, 2004.

Heckman, J. J., and J. A. Smith. “Assessing the Case for Social Experiments.” *Journal of Economic Perspectives* 9, no. 2 (1995): 85–110. doi:10.1257/jep.9.2.85.

Herley, C., and D. Florêncio. “Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy.” In *Economics of Information Security and Privacy*, edited by T. Moore, D. Pym, and C. Ioannidis, 33–53. Boston: Springer, 2010.

Holt, C., and R. Sherman. “Classroom Games: A Market for Lemons.” *Journal of Economic Perspectives* 13, no. 1 (1999): 205–214.

Houser, D., and J. Wooders. “Reputation in Auctions: Theory, and Evidence from eBay.” *Journal of Economics & Management Strategy* 15, no. 2 (2006): 353–369.

Huang, K., M. Siegel, K. Pearlson, and S. Madnick. “Casting the Dark Web in a New Light.” *MIT Sloan Management Review* 60, no. 4 (2019): 1–9.

Huck, S., H.-T. Normann, and J. Oechssler. “Two are Few and Four are Many: Number Effects in Experimental Oligopolies.” *Journal of Economic Behavior & Organization* 53, no. 4 (2004): 435–446.

Hutchings, A., R. Clayton, and R. Anderson. “Taking down Websites to Prevent Crime.” In *APWG Symposium on Electronic Crime Research (eCrime)*. Toronto: IEEE, 2016.

Hutchings, A., and T. Holt. “The Online Stolen Data Market: Disruption and Intervention Approaches.” *Global Crime* 18, no. 1 (2017): 11–30.

Jin, G. Z., and A. Kato. “Price, Quality, and Reputation: Evidence from an Online Field Experiment.” *The RAND Journal of Economics* 37, no. 4 (2006): 983–1004.

Kagel, J., and A. Roth. *The Handbook of Experimental Economics*. Princeton: Princeton University Press, 2016.

Kigerl, A. “Profiling Cybercriminals: Topic Model Clustering of Carding Forum Member Comment Histories.” *Social Science Computer Review* 36, no. 5 (2018): 591–609.

Ladegaard, I. “We Know Where You Are, What You are Doing and We Will Catch You: Testing Deterrence Theory in Digital Drug Markets.” *British Journal of Criminology* 58, no. 2 (2018): 414–433.

Leukfeldt, R., E. Kleemans, and W. Stol. “Origin, Growth and Criminal Capabilities of Cybercriminal Networks. An International Empirical Analysis.” *Crime, Law and Social Change* 67, no. 1 (2017): 39–53.

Li, L. “Reputation, Trust, and Rebates: How Online Auction Markets Can Improve Their Feedback Mechanisms.” *Journal of Economics & Management Strategy* 19, no. 2 (2010): 303–331.

Lusthaus, J. “Trust in the World of Cybercrime.” *Global Crime* 13, no. 2 (2012): 71–94.

Lusthaus, J. *Industry of Anonymity: Inside the Business of Cybercrime*. Cambridge: Harvard University Press, 2018.

Mell, A. *Promoting Market Failure: Fighting Crime with Asymmetric Information*. Oxford, 2015 Department of Economics Discussion Paper Series (Oxford, University of Oxford).

Melnik, M., and J. Alm. “Does a Seller’s Ecommerce Reputation Matter?” *The Journal of Industrial Economics* 50, no. 3 (2002): 337–349.

Paquet-Clouston, M. “Cybercrime Is Whose Responsibility? A Case Study of an Online Behaviour System in Crime.” *Global Crime* 19, no. 1 (2018): 1–21.

Paquet-Clouston, M., D. Decary-Hetu, and C. Morselli. “Assessing Market Competition and Vendors’ Size and Scope on AlphaBay.” *International Journal of Drug Policy* 54 (2018): 87–98.

Poulsen, K. *Kingpin*. New York: Crown Publishers, 2011.

Resnick, P., and R. Zeckhauser. “Trust among Strangers in Internet Transactions: Empirical Analysis of eBay’s Reputation System.” In *the Economics of the Internet and E-Commerce*, edited by M. Baye, 127–157. Amsterdam: Elsevier Science, 2002.

Rivlin, A. “Social Experiments: Their Uses and Limitations.” *Monthly Labor Review* 97, no. 6 (1974): 28.

Soska, K., and N. Christin. “Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem.” In *USENIX Security Symposium*. Washington, D.C: USENIX, 2015.

Van Buskirk, J., A. Roxburgh, M. Farrell, and L. Burns. “The Closure of the Silk Road: What Has This Meant for Online Drug Trading?” *Addiction* 109, no. 4 (2014): 517–518. doi:10.1111/add.12422.

Von Neumann, J., and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 2007.

Webster, M., and J. Sell. *Laboratory Experiments in the Social Sciences*. Amsterdam: Academic Press, 2007.

Yip, M., N. Shadbolt, and C. Webber. “Why Forums?: An Empirical Analysis into the Facilitating Factors of Carding Forums.” In *Web Science Conference*. Bloomington, IN: ACM, 2013.

Yip, M., C. Webber, and N. Shadbolt. “Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing.” *Policing and Society* 23, no. 4 (2013): 516–539.

Zhou, G., J. Zhuge, Y. Fan, K. Du, and S. Lu. “A Market in Dream: The Rapid Development of Anonymous Cybercrime.” *Mobile Networks and Applications* 25 (2020): 259–270.

Appendix 1 – Experiment instructions

The experiment instructions are presented below from the Control, Treatment 1, and Treatment 2.

All sections apply to all treatments, unless otherwise specified. Some sections apply only to Treatments 1 and 2, as their design included additional steps. For Treatment 1, these instructions are *italicised*. For Treatment 2, they are underlined.

Participants could only see instructions for their own treatment and were not informed which treatment they were participating in.

Welcome to this experiment and thank you for participating. Please read the following instructions carefully.

This is an experiment in decision-making in situations of uncertainty. It consists of four tasks, each of which will be explained as that part of the experiment begins. The amount of money you earn from this experiment may depend on the decisions you make, the decisions others make, and luck.

During the experiment, your earnings are given in tokens. At the end of the experiment you will be paid in CASH based on the following exchange rate:

25 tokens = 1 GBP

Other participants will not be able to see how much you have earned.

This experiment will last approximately 90 mins. Please do not talk or communicate with other participants during the experiment or look at other participants' computer screens. Please turn off your mobile phones to avoid any distractions.

If you have any questions, please raise your hand and someone will come to help you.

First task

In this task you are randomly paired up with two other participants one after the other.

In each pair, one of you is the FIRST MOVER and the other the SECOND MOVER. Each participant begins with **25 tokens**.

FIRST MOVERS choose to send none, some, or all of their 25 tokens to SECOND MOVERS. Once they have chosen the amount to send, SECOND MOVERS receive three times the amount sent to them. SECOND MOVERS then decide how many of their tokens to now send back to FIRST MOVERS, none, some, or all.

Neither the experimenters nor the participants know how much the FIRST MOVERS will send. As a result, SECOND MOVERS choose how much they would like to return to FIRST MOVERS for each possible amount they could have been sent.

Both FIRST and SECOND movers can only send multiples of five tokens (e.g. 0, 5, 10, 15, 20, 25) to one another.

For example:

- FIRST MOVER and SECOND MOVER both start with 25 tokens;
- FIRST MOVER sends 10 tokens to SECOND MOVER;
- SECOND MOVER receives 30 tokens and now has 55 tokens;
- SECOND MOVER sends 15 tokens back to FIRST MOVER;
- FIRST MOVER receives 15 tokens;
- FIRST MOVER now has a total payoff of: 25–10 + 15 = 30 tokens and SECOND MOVER of: 25 + 30–15 = 40 tokens.

Or:

- FIRST MOVER and SECOND MOVER both start with 25 tokens;
- FIRST MOVER sends 20 tokens to SECOND MOVER;
- SECOND MOVER receives 60 tokens and now has 85 tokens;
- SECOND MOVER sends 10 tokens back to FIRST MOVER;
- FIRST MOVER receives 10 tokens;

- FIRST MOVER now has a total payoff of: 25–20 + 10 = 15 tokens and SECOND MOVER of: 25 + 60–10 = 75 tokens.

You are either a FIRST MOVER or SECOND MOVER in each pair. Which mover you are in the first pair is decided at random. In the second pair you are the other mover.

You will be remunerated for one of these two decisions, chosen at random, either the one that you made as a FIRST MOVER or as a SECOND MOVER. Your payoff will depend on both the decision you made and the decision your partner made. You will not see the results of the decisions (and the corresponding payment) until the end of the experiment.

You will initially perform a trial round before the 'real' task begins, which will not impact your earnings.

Second task

In the second task, there are 30 rounds. You are either a BUYER or a VENDOR for the entirety of this task. In each round, you are randomly assigned to a group with two other participants. In each round, one group member is a BUYER and the other two are VENDORS.

At the beginning of each round, all VENDORS and BUYERS receive **50 tokens**.

Vendors begin by privately choosing a price (between 0 and 50 tokens) and a quality grade for their products – high, medium, or low; a higher grade costs more to produce and is worth more to buyers.

BUYERS then have a chance to purchase from either of the VENDORS in their group at the prices listed observing **price alone**. Otherwise, they can choose not to buy from either VENDOR.

After each sale, VENDORS are given the option not to 'send' the products to BUYERS, meaning they would receive money from the sale without paying for its production.

In SYBIL (Treatment 2): After each sale, VENDORS are given the option not to 'send' the products to BUYERS, meaning they would be receiving money from the sale without paying for its production. However, if a VENDOR chooses to 'send' the product, there is a 15% chance for each vendor during each round that this decision will be '**compromised'** and instead replaced by a decision 'not to send'. This does not affect the VENDOR's payoffs and they are charged for the production costs as they would have if the product had been 'sent'. Decisions 'not to send' always go through 'uncompromised'. Participants are not informed which 'sending' decisions have been 'compromised'.

BUYERS are then asked to rate the quality of the product they just acquired, based on its price, quality, and whether they received it or not, on a **scale of 0 to 5**. The average of all received ratings is displayed for the duration of the Task.

*In SLANDER (Treatment 1): BUYERS are then asked to rate the quality of the product they just acquired, based on its price, quality, and whether they have received them or not, on a **scale of 0 to 5**. These ratings build on each other throughout rounds and are available for the remainder of the experiment. However, there is a 15% chance for each BUYER during each round that these ratings get '**compromised'**, meaning they are discarded and replaced by a randomly chosen different rating. Every other rating than the one that was initially given by the BUYER has an equal chance to be chosen to replace it. Participants are not informed which ratings have been compromised and the compromised ratings are included in the VENDORS' averaged ratings.*

BUYERS can buy up to 1 unit of the commodity during a round. VENDORS can produce up to 1 unit in a round.

VENDORS in each group who are not chosen by buyers end the round with 0 tokens.

BUYERS who choose not to buy from either vendor end the round with 0 tokens.

For example (CONTROL):

- Round 4:
- VENDOR 1 (4.7/5 stars) chooses a price of 25 tokens and Medium quality for its product and VENDOR 2 (4.5/5 stars) chooses a price of 25 tokens and Low quality;
- BUYER can only see the prices on offer and chooses VENDOR 1's offer;

- VENDOR 1 then chooses to send the product to BUYER;
- BUYER rates the purchase 4 out of 5 stars.

Or:

- Round 17:
- VENDOR 1 (3/5 stars) chooses a price of 40 tokens and Medium quality for its product and VENDOR 2 (4.1/5 stars) chooses a price of 30 tokens and Low quality;
- BUYER can only see the prices on offer and chooses VENDOR 2's offer;
- VENDOR 2 then chooses not to send the product to BUYER;
- BUYER rates the purchase 0 out of 5 stars.

For example (TREATMENT 1):

- Round 4:
- VENDOR 1 (4.7/5 stars) chooses a price of 25 tokens and Medium quality for its product and VENDOR 2 (4.5/5 stars) chooses a price of 25 tokens and Low quality;
- BUYER can only see the prices on offer and chooses VENDOR 1's offer;
- VENDOR 1 then chooses to send the product to BUYER;
- BUYER rates the purchase 4 out of 5 stars;
- *The rating is not compromised and stays 4 out of 5 stars.*

Or:

- Round 17:
- VENDOR 1 (3/5 stars) chooses a price of 40 tokens and Medium quality for its product and VENDOR 2 (4.1/5 stars) chooses a price of 30 tokens and Low quality;
- BUYER can only see the prices on offer and chooses VENDOR 2's offer;
- VENDOR 2 then chooses not to send the product to BUYER;
- BUYER rates the purchase 0 out of 5 stars;
- *The rating is compromised and becomes 3 out of 5 stars.*

For example (TREATMENT 2):

- Round 4:
- VENDOR 1 (4.7/5 stars) chooses a price of 25 tokens and Medium quality for its product and VENDOR 2 (4.5/5 stars) chooses a price of 25 tokens and Low quality;
- BUYER can only see the prices on offer and chooses VENDOR 1's offer;
- VENDOR 1 then chooses to send the product to BUYER;
- The sending decision is compromised to 'not to send';
- BUYER rates the purchase 4 out of 5 stars.

Or:

- Round 17:
- VENDOR 1 (3/5 stars) chooses a price of 40 tokens and Medium quality for its product and VENDOR 2 (4.1/5 stars) chooses a price of 30 tokens and Low quality;
- BUYER can only see the prices on offer and chooses VENDOR 2's offer;
- VENDOR 2 then chooses not to send the product to BUYER;
- BUYER rates the purchase 0 out of 5 stars.

You will be remunerated for 5 of the 30 rounds in this Task, all chosen at random.

You will initially perform a trial round before the 'real' 30 rounds begin, which will not impact your earnings.

Third task

In this task you are randomly paired with two other participants in the session and perform two decisions, one as FIRST MOVER and the other as SECOND MOVER, each in a different pair, like in the First task. You will be remunerated for one of these two decisions chosen at random, either the one that you made as a FIRST MOVER or as a SECOND MOVER.

Each participant begins with **25 tokens**.

FIRST MOVERS choose to send none, some, or all of their 25 tokens to SECOND MOVERS. Once they have chosen the amount to send, SECOND MOVERS receive three times the amount sent to them. SECOND MOVERS then decide how many of their tokens to send back to FIRST MOVERS, none, some, or all.

Neither the experimenters nor the subjects know how much the FIRST MOVER will send. As a result, SECOND MOVERS choose how much they would like to return for each possible amount they could be sent.

Both FIRST and SECOND movers can only send multiples of five tokens (e.g. 0, 5, 10, 15, 20, 25) to one another.

Fourth task

You will then be asked to complete several survey questions following this study. Some questions are demographic (e.g. age, gender, occupation, degree), while others are about your strategy in the experiment and perception of other players' strategies. Your answers will not be matched with your name and will be kept anonymous, so please answer honestly. This survey takes approximately 5 minutes to complete. Please take your time to answer all of the questions while we prepare your individual cash payment.

You will then find out the outcome and your total payoffs from the experiment. All final payments will be rounded up to the nearest Pound.

Please stay seated until the experimenter calls you to receive your payment.

Thank you for your participation!

Appendix 2 – Vendor and buyer payoffs

N.B. Payoff information was provided on a separate sheet based on the participants' assigned roles. Each participant only received one bit of information, either that of a vendor or a buyer.

In this Task you are a **VENDOR**.

The table below shows production costs for different grades, buyers cannot see this information:

Grade	High	Medium	Low
Production cost to vendors	30	20	10

The period payoff for VENDORS is: **50 tokens + vendor's price – cost of the grade produced (if sent)**

VENDORS in each group who are not chosen by buyers end the round with 0 tokens.

In this Task you are a **BUYER**.

The table below shows values for different grades, vendors cannot see this information:

Grade	High	Medium	Low
Value to buyer	45	30	15

The period payoff for BUYERS is: **50 tokens + value of the grade purchased (if sent) – vendor's price**

BUYERS who choose not to buy from either vendor end the round with 0 tokens.