European Sociological Review, 2017, Vol. 33, No. 6, 752–764
doi: 10.1093/esr/jcx072
Advance Access Publication Date: 17 October 2017
Original Article

Order without Law: Reputation Promotes Cooperation in a Cryptomarket for Illegal Drugs

Wojtek Przepiorka[1,2,*], Lukas Norbutas[1,3] and Rense Corten[1]

[1]Department of Sociology / ICS, Utrecht University, Padualaan 14, Utrecht, 3584 CH, The Netherlands, [2]Nuffield College, New Road, Oxford OX1 1NF, UK and [3]Netherlands Institute for the Study of Crime and Law Enforcement, De Boelelaan 1077a, Amsterdam 1081 HV, The Netherlands

*Corresponding author. E-mail: w.przepiorka@uu.nl

Submitted March 2017; revised July 2017; accepted September 2017

Abstract

The emergence of large-scale cooperation in humans poses a major puzzle for the social and behavioural sciences. Reputation formation—individuals' ability to share information about others' deeds and misdeeds—has been found to promote cooperation. However, these findings are mostly based on small-scale laboratory and field experiments or on data gathered from online markets embedded in functioning legal systems. Using a unique data set of transactions in a cryptomarket for illegal drugs, we analyse the effect of buyers' ratings of finished transactions on sellers' business success. Cryptomarkets are online marketplaces in the so-called Dark Web, which can only be accessed by means of encryption software that conceals users' identities and locations. The encryption technology makes it virtually impossible for law enforcement to intervene in these market exchanges. We find that sellers with a better rating history charge higher prices and sell their merchandise faster than sellers with no or a bad rating history. Our results demonstrate how reputation creates real incentives for cooperative behaviour at a large scale, in the absence of law enforcement and among anonymous actors with doubtful intentions. Our results thus challenge the institutional and social embeddedness of actors as necessary preconditions for the emergence of social order in markets.

Introduction

Humans' ability to overcome individual self-interest to create a larger benefit for the collective has received considerable attention in the social and behavioural sciences in the past three decades (Axelrod, 1984; Nowak, 2006; Bowles and Gintis, 2011). A simple mechanism that has been shown to promote cooperation in humans is our ability to share information about our peers' deeds and misdeeds with third parties. Such information sharing contributes to the formation of individuals' reputations (Dunbar, 2004; Sommerfeld *et al.*, 2007; Feinberg, Willer and Schultz, 2014). Individuals with a reputation to lose have a strong incentive to behave cooperatively and are therefore attractive partners in social and economic exchange (Shapiro, 1983; Kollock, 1994; Sylwester and Roberts, 2013; Milinski, 2016). More generally, it has been argued that reputational incentives provide a more efficient means to uphold norm compliance and order in society than other forms of sanctioning (Ellickson, 1991; Milinski, Semmann and Krambeck, 2002; Willer, 2009; Grimalda, Pondorfer and Tracer, 2016; Wu, Balliet and van Lange, 2016).

Throughout human history, reputation mechanisms have also facilitated mutually beneficial economic exchanges. In the absence of formal institutions protecting property rights and enforcing contractual commitments, the transfer of reputation information through dense social networks has promoted good business conduct (Hillmann, 2013). For example, Greif (1989) describes how Maghribi traders in medieval Europe organized in coalitions in which they exchanged information about their agents' reputations to reduce the trust problems arising in long-distance trade. Hillmann and Aven (2011) describe the role reputation played in the development of corporate capitalism in Russia around the turn of the nineteenth century. At that time, entrepreneurs faced the trade-off of limiting their interactions to local partners to ensure compliance and interacting with partners beyond their social networks offering more profitable but riskier businesses.

These historical examples illustrate how the lack of formal institutions promoting economic exchange is replaced by informal institutions at work in social groups in which information about actors' deeds and misdeeds is exchanged and selective incentives upheld (Nee 2005). However, there are several historical accounts of centralized reputation systems which facilitated economic exchange without requiring actors to be closely connected via a social network. Also in the early middle ages, the Champagne Fairs in France became a meeting point for traders from all over Europe. Promoted by the use of bookkeeping and cashless payment, a private adjudication system evolved that allowed tracking fraudulent traders and excluding them from future fairs (Milgrom *et al.*, 1990). In the late nineteenth century, so-called credit bureaus started to emerge, which collected and shared information about borrowers' credit histories creating reputational incentives for timely debt repayment (Jappelli and Pagano, 2002; Carruthers, 2013). These examples are prototypical for the centralized reputation systems that are today's standard for governing online market exchanges (Dellarocas, 2003; Diekmann and Przepiorka, forthcoming).

In online markets such as eBay, thousands of anonymous buyers and sellers trade with each other every day across large geographic distances. Via an electronic feedback system, traders can comment on each other's conduct after finished transactions with positive or negative ratings and short text messages, and these ratings constitute online traders' reputations. The advent of the Internet and the emergence of online markets have created ample opportunities to study the effectiveness of reputation systems to promote cooperation among anonymous traders at a large scale (Kollock, 1999; Resnick and Zeckhauser, 2002; Dellarocas, 2003; Bolton, Greiner and Ockenfels, 2013; Diekmann *et al.*, 2014). In particular, it has been shown how electronic reputation systems create incentives for traders' cooperative behaviour without requiring these traders to be embedded in social networks (Granovetter, 1992; Diekmann *et al.*, 2014). However, the working of the reputation mechanism has thus far only been established under favourable conditions. The majority of online markets are embedded in functioning legal systems attracting and backing up trades among individuals with mostly good intentions. It is thus an open question whether reputation formation net of legal and moral assurances is sufficient to promote cooperation in a large group of strangers.

Here we study the functioning of the reputation mechanism in a cryptomarket for illegal drugs. Cryptomarkets are online marketplaces in the Dark Web, which can only be accessed by means of encryption software that conceals users' identities and locations (Martin, 2014). Trades in cryptomarkets include forged personal documents, hacked user accounts, weapons, etc., with illegal drugs constituting the largest proportion of trades (Christin, 2012; Soska and Christin, 2015). Globally, almost 10 per cent of drug users reported ever buying drugs from cryptomarkets (Global Drug Survey, 2016). The encryption technology makes it virtually impossible for law enforcement to intervene. Hence, given the lack of legal deterrent, traders' good intentions are highly uncertain at best. This creates severe trust problems between buyers and sellers, as it makes buyers vulnerable to sellers' fraudulent transactions (Dasgupta, 1988; Coleman, 1990). However, in the same way as online markets for licit goods, cryptomarkets use electronic reputation systems which allow buyers to rate sellers after finished transactions (Bartlett 2014; Hardy and Norgaard, 2016). This makes cryptomarkets the ideal settings to test the potential of the reputation mechanism to bring about 'order without law' (Ellickson, 1991).[1]

In the next section, we describe the reputation mechanism, previous approaches to its studying, the set-up of the cryptomarket that we study, and state our hypotheses. In the 'Data and Methods' section, we describe the data and data gathering process, the variables we used in our analyses, and our model estimations. Thereafter we present our results and, in the final section, we conclude with a discussion of our findings.

Reputation in Markets and the Problem of Embeddedness

Acquiring a good seller reputation in online (and offline) markets is costly because it can only be achieved through good business conduct over an extended period of time. Sellers who lack trustworthy-making properties, such as long-term business interests or honest intentions, will not bother to enter the market and build a good reputation by behaving cooperatively. Hence, based on sellers' reputations, buyers can infer these sellers' trustworthiness and choose the sellers they prefer to buy from. However, trustworthy sellers who enter the market and, therefore, have not yet built their reputation, are indistinguishable from their untrustworthy competitors. New, trustworthy sellers must therefore allow prices to compensate potential buyers for the risk they take when trading with 'unknown' sellers. Once these sellers have built their reputations, they can charge higher prices, which will compensate them for their initial investment in reputation (Shapiro, 1983; Friedman and Resnick, 2001; Przepiorka, 2013).[2] From this reasoning it follows that sellers with a better reputation will achieve higher prices because buyers are willing to pay higher prices for these sellers' products.

This conjecture has been corroborated in more than two dozen studies, most of which analyse the effect of sellers' reputations on the probability of product sale and selling price using eBay auction data (for a review, see Diekmann *et al.*, 2014). Although the costliness of building a good reputation constitutes an important deterrent for fraudulent sellers, the legal system in which online markets are embedded deters fraud and promotes trust and large-scale cooperation in its own right (Diekmann and Przepiorka forthcoming; Fligstein, 2001; Güth and Ockenfels, 2003; Pavlou and Gefen, 2004; Beckert, 2009; Bakos and Dellarocas, 2011). First, the legal system maintains a non-negligible threat that fraudulent business conduct will be prosecuted and punished. Secondly, online market platforms can be made accountable for sellers' misconduct by their community of buyers, who can easily turn to alternative platforms. As a consequence, platform providers have a strong incentive to protect buyers from fraud by, for example, monitoring sellers' activities and sanctioning bad behaviour (e.g. by banning sellers). However, unlike online markets for licit goods, cryptomarkets cannot work with law enforcement to combat fraudulent behaviour (Calkins *et al.*, 2008). Thirdly, buyers are insured against fraud to a certain extent if they use credit card payment. Although such insurances do not eliminate the trust problem, they reduce the material losses buyers may expect when trading in online markets. In sum, legal assurances preselect sellers with good intentions, incentivize platform providers to enforce cooperation, and reduce the risk of large monetary losses. As a consequence, buyers will have high *a priori* expectations as to online sellers' trustworthiness (Güth and Ockenfels, 2003; Lindenberg, 2017).[3]

In the light of these considerations, it is an open question whether the reputation mechanism, as instituted in many electronic rating systems, promotes trust and cooperation in online markets net of their embeddedness in well-functioning legal systems. Backed by legal and moral assurances, reputation may function as a mere coordination device, which facilitates buyers' choices among the plethora of sellers offering the same products, rather than solving cooperation problems (Beckert, 2009; Frey and van de Rijt, 2016; Przepiorka and Aksoy, 2017). More importantly, in the absence of legal and moral assurances, reputation systems may fail to attract a critical mass of traders because they may be regarded as an insufficient safeguard of mutually beneficial economic exchange.

One way to address this question is to study the functioning of the reputation mechanism in an extra-legal context, for example, as we do, in a cryptomarket for illegal goods. If we find reputation effects even in the absence of legal and social conditions that deter opportunistic actors and promote trust, this would strongly reinforce the idea that reputation systems enable the bottom-up emergence of cooperation in large groups of self-regarding actors. We thus re-evaluate the claim from earlier research that reputation affects market outcomes of sellers in the context of cryptomarkets for illegal drugs, and derive hypotheses for our specific study context, the Cryptomarket Silk Road 1.0.[4]

The Cryptomarket Silk Road 1.0

We use data from the first cryptomarket, Silk Road 1.0 (Christin, 2012), to study in how far sellers reap the benefits of a good reputation and in how far buyers take into account sellers' reputations when deciding which seller to buy from. Silk Road 1.0 started operating in February, 2011 and was closed after its owner was arrested in October 2013 (Barratt and Aldridge, 2016). Our data contain information on all item listings that were online between 3 February 2012 and 24 July 2012, including item names and descriptions, categories, prices, and item-specific feedback messages. Since each item listing contains an encrypted vendor identifier, feedback can be attributed to individual sellers who were active during this period.

A typical transaction on Silk Road is initiated by the seller, who decides on the number of items and the item price of his or her product, and posts the offer online. Buyers can then buy the item at the specified price as long it is available. A buyer first sends the money to an escrow service, which releases the money and transfers it to the seller when the buyer confirms receipt of the item. Although the escrow service mitigates the trust problem to a certain extent as it protects the buyer from spending the money without receiving anything in return, the trust problem remains with regard to the quality of the product and the competence of the seller. Buyers cannot withhold payment because of quality reasons, as this would require using or testing the product. Moreover, there is more to a transaction than the money and the product. A seller's good reputation also stands for his or her professional handling of the transactions. Since buyers have to provide a postal address the seller can send the product to, buyers have to trust the seller to wrap the product inconspicuously, send it to the right address, and maybe provide a refund should the product get lost on its way to the buyer (Bartlett 2014: Ch. 5).

In Silk Road 1.0, a finished transaction receives the highest rating (i.e. five stars) by default. Buyers can change the default rating to a four-, three-, two-, or one-star rating and add a text comment. The sum of these seller ratings establishes a seller's reputation in the market. The information about a seller's reputation is conspicuously displayed and can be considered by buyers deciding which seller to buy from. We use the sum of a seller's five-star ratings and non-five-star ratings as an indicator of this seller's reputation, and keep track of sellers' rating history over time (see 'Data and Methods' section). We use item prices, which are set by the sellers, as an indicator for sellers' cashing in the premium for their good reputation or giving a discount in case of no or a bad rating history. We use the speed at which items are sold, which is determined by buyers, as an indicator for the trust buyers have in sellers with a certain rating history. Based on the theoretical argument in the first paragraph of the previous section and the set-up of Silk Road 1.0, we can formulate the following four hypotheses:

H1: The more five-star ratings a seller has, the more he or she will charge for his or her items.
H2: The more non-five-star ratings a seller has, the less he or she will charge for his or her items.
H3: The more five-star ratings a seller has, the more items he or she will sell per day.
H4: The more non-five-star ratings a seller has, the less items he or she will sell per day.

Data and Methods

Data

We use a data set containing 24,385 items collected by Christin (2012) on Silk Road 1.0 between 3 February 2012 and 24 July 2012. To avoid bias due to unobserved item heterogeneity, we select a subset of illegal drug items for our analyses (Diekmann *et al.*, 2014). Some types of drugs (e.g. LSD) are sold in different forms (e.g. pills, powder, blotter), which vary in weight and substance concentration, making the calculation and comparability of item prices per gram more difficult. Our subset comprises seven categories of illegal drugs: weed, hash, cocaine, ketamine, MDMA, heroin, and meth. We limit our analyses to these categories because of their size in terms of the number of item listings and the comparability of items within each category. Item listings in these categories account for 24.6 per cent (6,005) of all item listings and 37.3 per cent of all feedback messages. To this sample we add 211 items listed in the general categories 'Drugs' and 'Cannabis', which we could identify as also belonging to one of the seven categories specified above. This results in 6,216 items.

Information on item weight is not available in a standardized form in the original data set. Item weight information is only provided as part of the item name or item description. We extract item weight from the item name or description manually and exclude items which have no information on weight or have weight information not comparable with the majority of items in the category (e.g. pre-rolled joints of marijuana in category 'weed', where the majority of items are sold in grams of loose marijuana). At this stage we exclude 430 items. Moreover, we exclude 111 items that are given away (e.g. as 'freebies', 'samples', or 'lotteries') or are included as custom listings for a specific buyer. This leaves us with 5,675 items offered by 550 different sellers.

Of these 5,675 items, 2,522 (44.4 per cent) received no feedback messages during the time of data collection. Since we use the number of feedback messages an item received in total as a measure for the number of item sales (see below), we have no evidence that these items generated any sales. Note that 67 per cent of these items were online for 5 days or less, whereas only 11 per cent of the items for which at least one sale was recorded were online 5 days or less. In other words, a large majority of the items without recorded sales were online for a relatively short time. We cannot be sure that the sellers posting these 2,522 items had the intention to sell them as specified. These sellers may have offered these items but not reacted to any buyer requests or taken the items offline before completing any transactions. We therefore exclude these items from our main analyses. Our main analyses are based on 3,153 items, for each of which at least one sale was recorded. These items were offered by 445 different sellers. However, in the online Supplementary Material we also discuss the results of logistic regressions that are based on all 5,675 items and estimate the probability that at least one sale was recorded conditional on covariates. The results of this analysis are consistent with the results reported here.

Variables and Model Estimations

The unit of analysis is an item offered by a seller. Our two target variables are item price per gram (in USD) and number of item sales per day. We calculate item price per gram by dividing item price by item weight. Since no direct information on the number of sales per item is available in Silk Road 1.0, we derive the number of item sales from the total number of ratings a seller received for a particular item. Since a five-star rating is automatically awarded after a finished transaction, this approach provides an accurate estimate of the number of item sales (Soska and Christin, 2015; Décary-Hétu, Paquet-Clouston and Aldridge, 2016). To obtain the number of item sales per day, we divide the number of item sales by the time in days the item was observed online.

To test Hypotheses 1 and 2, we use a regression model with the log-transformed item price per gram (in USD) as the dependent variable; to test Hypotheses 3 and 4, we use a regression model with the log-transformed number of sales per day as dependent variable. In both models, we use the same explanatory and control variables with two exceptions. In our model of item price, we use the log number of total item sales as a measure for the quantity of an item initially in the seller's possession. Recall that item prices are set by the sellers. Since sellers are likely to pass on quantity discounts to their customers, the quantity of an item in the seller's possession will negatively affect item price. For obvious reasons, we do not use the log number of item sales as an explanatory variable in our models of sales per day, but we use log item price per gram for it will negatively affect item sales per day.

Our main explanatory variables are sellers' positive and negative reputation scores. We calculate a seller's reputation scores by summing up the number of five-star ratings and non-five-star ratings the seller has received by the time his or her item is first listed online. Note that we aggregate the five-star and non-five-star ratings across all items of a seller, also the items not included in our analyses. This is done to replicate the aggregated seller reputation score displayed in Silk Road 1.0 as closely as possible. We use the log number of five-star ratings as a measure of positive seller reputation and the log number of non-five-star ratings as a measure of negative seller reputation. Five-star ratings account for 95.8 per cent of all ratings in our sample; non-five-star ratings are given in extraordinary cases and might have negative impact on a seller's reputation irrespective of the actual number of stars.

We include several control variables in our models. To reduce the number of control variables, we pool the seven categories of illegal drugs in three price categories: Low price (weed and hash), medium price (cocaine, ketamine, and MDMA), and high price (heroin and meth). We use a set of three dummy variables to account for the three different categories with low price items as the reference category. Items in Silk Road 1.0 are offered in quantities ranging from 0.05 g to 1,000 g (see Table 1). We use log-transformed item weight (in grams), which we expect to have a negative effect on item price because of quantity discounts offered by sellers. Item weight will also have a negative effect on the number of sales per day because of a lower demand for bulk offers (Aldridge and Décary-Hétu, 2016). We also use dummy variables for shipping locations of each item as offered by the sellers. Since sellers who ship their items internationally face additional risks (Décary-Hétu *et al.*, 2016), these sellers might charge a price premium. For most sellers we have information on their country of origin and the countries and regions these sellers ship their items to. Based on this information, we distinguish between sellers who only ship domestically (i.e. within their country of origin), sellers who also ship their items abroad and sellers with unknown shipping preferences. Sellers who only ship domestically constitute the reference category. One dummy variable accounts for items of poor quality, an attribute sometimes specified in the titles of low price items only. Finally, since the data are right censored, we include a dummy variable to mark items that were listed online on the last 2 days of data collection. Table 1 contains the descriptive statistics of the main variables used in our analyses.

We do not have information on sellers' ratings from before the start of data collection. We therefore calculate sellers' reputations by summing up the number of five-star ratings and non-five-star ratings these sellers receive throughout the observation period. Since we perform our data analysis at the item level, we use the sum of ratings a seller received by the time an item was first listed in the market as an indicator of the seller's reputation (also see above). In other words, our data allow us to capture the change in sellers' reputations over time; our ability to compare sellers based on their number of five-star and non-five-star ratings is rather limited. We therefore estimate regression models with seller fixed effects (FE), which are solely based on the within-seller variation in dependent and explanatory variables (Allison, 2009). These models estimate how changes in seller's reputation, compared to seller's average

Table 1. The table lists descriptive statistics of the main variables used in our analyses

Variable name	N	Mean	SD	Median	Minimum	Maximum
Item sales and duration online						
# item sales	3,153	20.72	58.94	5	1	1501
item online in days	3,153	50.44	56.17	28	0.5	382
# item sales per day	3,153	0.45	0.69	0.25	0.01	10.83
Seller ratings at time item was first seen						
# five-star ratings	3,153	148.3	279.9	43	0	2615
# non-five-star ratings	3,153	4.89	12.36	0	0	149
Low-price products (weed, hash)						
weight in g	2,297	18.15	63.57	5	0.25	1000
price in USD per gram	2,297	15.50	7.30	14.61	1.46	115.8
Medium-price products (ketamine, MDMA, cocaine)						
weight in g	562	7.17	45.00	1	0.05	1000
price in USD per gram	562	92.26	57.52	80.58	8.41	464.1
High-price products (meth, heroin)						
weight in g	294	1.40	3.87	0.5	0.10	56
price in USD per gram	294	217.6	140.4	173.8	33.72	992.8

Note: The data comprise $N = 3153$ items which were online for 50 days and generated 21 sales on average. Only items for which at least one transaction was recorded are considered in this analysis (see 'Data and Methods' section). For each item, we calculated the number of five-star and non-five-star ratings the seller of an item had received up to the time point the item was first seen online. The seller of an item has 148 five-star and 5 non-five-star ratings on average. We divided items in three price categories. Low-price items ($N = 2297$) comprise weed and hash and are sold for USD 16 per gram in packages of 18 g on average. Medium-price items ($N = 562$) comprise ketamine, MDMA, and cocaine, and are sold for USD 92 per gram in packages of 7 g on average. High-price items ($N = 294$) comprise meth and heroin, and are sold for USD 218 per gram in packages of 1 g on average.

reputation across all observations, affect changes in seller's item prices and number of sales.

Results

Table 2 shows the regression model estimations with seller FE. In both the model of item price (M1) and the model of sales per day (M2), the coefficient estimates of the log number of five-star ratings are positive and the coefficient estimates of the log number of non-five-star ratings are negative. The four coefficients are statistically significant providing first evidence in support of our hypotheses.

Recall from Table 1 that the range in the number of five-star ratings in our data is considerably larger than the range in the number of non-five-star ratings. In what follows, we use 10-fold increases in five-star ratings (e.g. 50 vs. 500) and 3-fold increases in non-five-star ratings (e.g. 7 vs. 21) to calculate the effect of sellers' rating histories on item price and selling speed.

In Model M1, if the number of five-star ratings increases by a factor 10, sellers increase the item price by $100 \times [\exp(0.029 \times \ln(10)) - 1] = 6.8$ per cent, and if the number of non-five-star ratings increases by a factor 3, sellers decrease the item price by $100 \times [\exp(-0.044 \times \ln(3)) - 1] = (-)\ 4.7$ per cent. Based on the average selling price of a medium-price item of USD 92.26 (see Table 1), these changes correspond to USD 6.29 and USD −4.35, respectively. These results clearly support our first two hypotheses (*H1* and *H2*) and show that sellers' price adjustments in response to changes in their reputation can be substantial. These results are visualized in Figure 1A.

The coefficients of the control variables point in the expected directions. The higher its product category and the higher its quality, the higher is the price the seller asks for the item. The negative and statistically significant coefficients of the log number of sales and item weight both indicate that sellers give quantity discounts to buyers. Although sellers who also ship their items abroad tend to charge higher prices than sellers who only ship domestically, the difference is statistically insignificant.

Setting the price of an item is a deliberate choice made by sellers. Therefore, changes in item price do not directly tell us whether buyers infer trustworthiness from sellers' reputations and act accordingly. Given sellers' reputations and item characteristics, it is mainly the buyers who determine how quickly an item sells. Selling speed thus constitutes a better measure for how sellers' reputations help buyers to overcome the trust problem.

In Model M2, we find similar results in terms of item sales per day, as we found in our model for item price (M1). The coefficient estimates of the log number of five-star ratings and log number of non-five-star ratings

Table 2. Regression models of item price and sales per day with seller FE

Variable name	Log(item price per gram in USD) M1	Log(# item sales per day) M2
Const.	2.974***	1.402***
	(0.036)	(0.259)
Item variables		
Log(item price per gram in USD)		−0.823***
		(0.089)
Log(weight in gram)	−0.208***	−0.502***
	(0.009)	(0.079)
Low price	(reference)	(reference)
Medium price	1.506***	0.908***
	(0.065)	(0.218)
High price	2.035***	1.271***
	(0.095)	(0.310)
Poor quality (weed and hash)	−0.756***	0.190
	(0.078)	(0.196)
Last 2 days	0.114***	0.174*
	(0.027)	(0.077)
Seller variables		
Log(# five-star ratings + 1)	0.029***	0.060*
	(0.007)	(0.028)
Log(# non-five-star ratings + 1)	−0.044**	−0.188***
	(0.016)	(0.057)
Log(# item sales)	−0.035***	
	(0.006)	
Seller ships to		
Unknown	−0.012	−0.006
	(0.064)	(0.207)
Domestic only	(reference)	(reference)
Foreign	0.035	0.063
	(0.036)	(0.130)
N_1	3153	3153
N_2	445	445
Adjusted R^2	0.77	0.22

Notes: The table lists coefficient estimates and cluster–robust standard errors (***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, for two-sided tests) of FE regression models. The target variable in Model M1 is the log-transformed item price per gram in USD. The target variable in Model M2 is the log-transformed number of item sales per day. N_1 denotes the number of cases (items), and N_2 denotes the number of clusters (sellers). Both models include seller FE. Figures 1 A and 1B are based on Models M1 and M2, respectively.

can be interpreted as follows. If the number of five-star ratings increases by a factor 10, item sales per day increase by $100 \times [\exp(0.060 \times \ln(10)) - 1] = 14.7$ per cent, and if the number of non-five-star ratings increases by a factor 3, item sales per day decrease by $100 \times [\exp(-0.188 \times \ln(3)) - 1] = (-)18.7$ per cent. Based on these changes, a median seller, who sells one item in 4 days (see Table 1), would need 3.5 or 5 days, respectively, to sell that item. These results clearly support hypotheses *H3* and *H4*. These results are visualized in Figure 1B. The control variables exhibit the same effects as in our model of item price. Unsurprisingly, item price has a significantly negative effect on the number of sales.

The effects of sellers' reputation on selling speed appear relatively small. To a large extent, this is a result of our accounting for differences in unobserved, time-constant seller characteristics in our model estimations (Allison, 2009). However, buyers choosing a seller compare offers from different sellers (Snijders and Weesie, 2009). Therefore, we also estimated ordinary least square regression models, which take between-seller variability of seller and item characteristics into account (see online Supplementary Material). Based on these estimations, the effects of five-star ratings on selling speed are up to three times larger than the effects shown in Figure 1B, but these effects can only partly be interpreted causally.

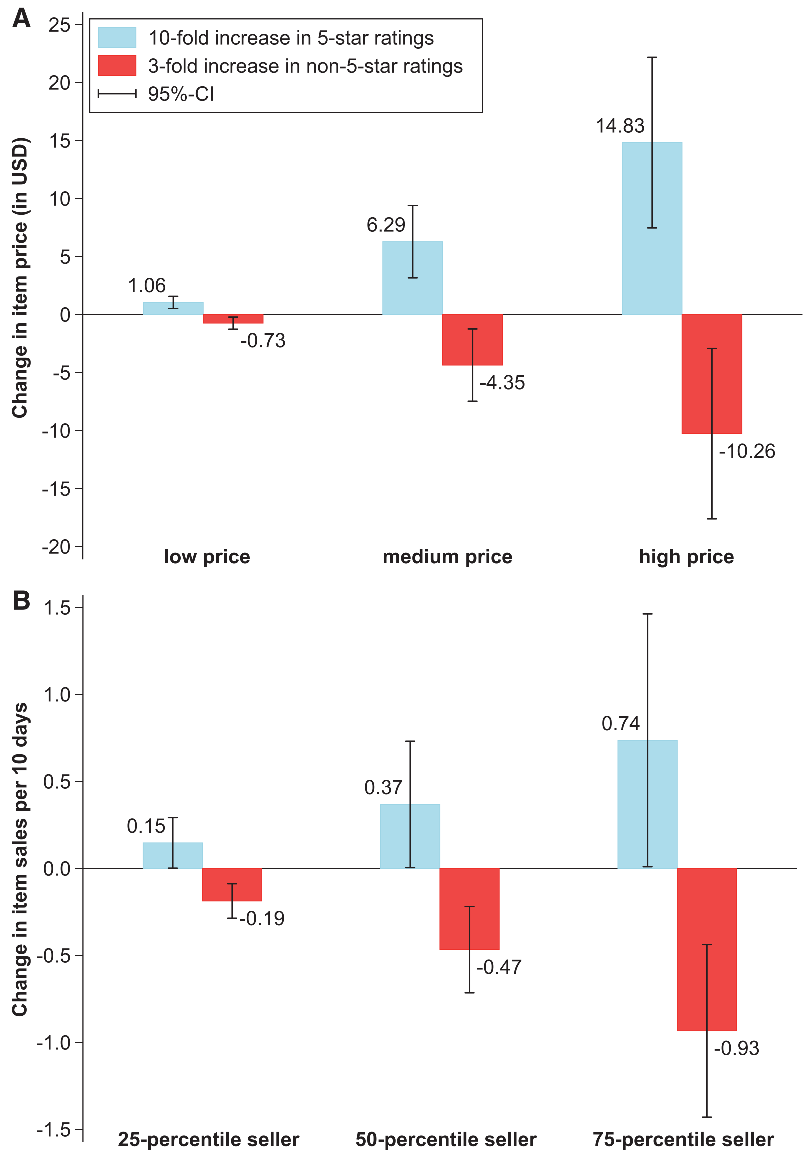

Figure 1. Shows the changes in item price (**A**) and selling speed (**B**) due to changes in the number of five-star and non-five-star ratings of a seller. In (**A**), changes in item price are calculated relative to the average price of a low-price, medium-price, and high-price product; in (**B**) changes in selling speed are calculated relative to sellers at the 25th, 50th, and 75th percentile in terms of selling speed. Due to a 10-fold increase in the number of five-star ratings, a seller increases the item price by USD 1.06, USD 6.29, and USD 14.83 for a low-price, medium-price, and high-price product, respectively. Correspondingly, due to a 3-fold increase in the number of non-five-star ratings, a seller decreases the item price by USD 0.73, USD 4.35, and USD 10.26 for a low-price, medium-price, and high-price product, respectively. These results support hypotheses *H1* and *H2*: sellers reap the benefits of a good reputation by increasing prices and give a discount if they have not yet established a good reputation or their reputation decreased due to non-five-star ratings. Due to a 10-fold increase in the number of five-star ratings, a median seller sells 0.37 items more in 10 days, and due to a 3-fold increase in the number of non-five-star ratings the median seller sells 0.47 items less in 10 days. These effects are larger for 75th percentile sellers, who have a higher frequency of sales, and smaller for 25th percentile sellers, who have a lower frequency of sales than a median seller. These results support hypotheses *H3* and *H4*: Buyers are more eager to buy from sellers with a good reputation and more reluctant to buy from sellers who have not yet established a good reputation or who have received non-five-star ratings in the past.

Discussion

We study how cooperation is established between anonymous individuals in a cryptomarket for illegal goods. We use longitudinal data on market transactions from the first major cryptomarket, called 'Silk Road', to test whether reputation formation can promote cooperation between buyers and sellers in an environment of high uncertainty and in the absence of a centralized authority. We use all available market transactions in the seven largest categories of illegal drugs to test whether high (low) buyer ratings increase (harm) sellers' market success in terms of pricing and sales. We find that sellers' rating histories affect the behaviour of both sellers and buyers. Sellers react to changes in their reputation by adjusting the prices of their goods. Well-reputed sellers reap market benefits by increasing prices, while sellers with lower reputations decrease their prices to compensate potential buyers for the risk they take when buying from them (Shapiro, 1983; Friedman and Resnick, 2001; Przepiorka, 2013). We also find that sellers with better reputations sell more goods over the same period of time. Although we do not observe buyers' choices of particular sellers, the higher selling speed of well-reputed sellers' items suggests that buyers choose sellers based on these sellers' rating histories. Finally, we find that negative ratings have a larger absolute effect on sellers' prices and sales than positive ratings. Negative asymmetry, or a large impact of negative information about partner's trustworthiness on withholding trust, has been observed previously in online markets and experimental settings (Standifird, 2001; Bozoyan and Vogt, 2016). Our findings suggest that in cryptomarkets too damaged reputations are hard to repair (Matzat and Snijders, 2012).

Our research contributes to the agenda of the new institutionalism in economic sociology (Nee 2005; Beckert, 2009; Hillmann, 2013) in at least two ways. First, we show that reputation formation is a robust mechanism to foster trust and cooperation in online markets net of legal assurances, verifiable identities, or a positive self-selection of mostly law-abiding citizens. In online markets for licit goods, the trust problem inherent in economic exchanges is mitigated but not entirely dissolved by legal and moral assurances. Our results thus corroborate that reputation systems in general can be an essential organizational assurance which, if well-designed, protect online traders from being cheated or in other ways dissatisfied by their peers. Secondly, in a historical perspective, cryptomarkets constitute a next phase in the evolution of market institutions. In our article, we describe a case that illustrates how cooperative market exchanges are possible at a large scale in the absence of formal institutions established by nation states and informal institutions at work in small social groups (Nee 2005). In fact, formal and informal institutions such as legal systems and social norms, respectively, have not disappeared, but they have lost their influence in the governing of online market exchanges (Przepiorka and Aksoy 2017).

Our findings also illustrate more generally the potential of data generated by the hidden corners of the World Wide Web for studying fundamental social processes. Since the shutdown of Silk Road 1.0, many new cryptomarkets have emerged, which have developed a wide range of institutional arrangements that promote trust between buyers and sellers, but also online traders' trust in cryptomarket platforms. Traders' trust in market platforms has become essential for establishing cooperation on the Dark Web, since absence of law enforcement and full anonymity also brought uncertainty with regard to trustworthiness of cryptomarkets as institutions. After several cases of large-scale fraud committed by owners of major cryptomarkets, more recent marketplaces signal their trustworthiness to potential traders by investing in technical innovations. Such innovations include, for example, multi-signature escrow systems that prevent marketplaces from abusing funds in the escrow system (Bartlett 2014: Ch. 5; Barratt and Aldridge, 2016), more complex website code bases that require cryptomarket owners to have higher levels of technical knowledge (Branwen, 2016) or finances to start and maintain the marketplace, but also in social innovations. The latter is exemplified by some cryptomarkets' introduction of hierarchical systems among buyers and sellers, where only well-reputed buyers can gain access to goods of well-reputed sellers. In other words, only after having engaged in some illicit transactions are buyers trusted with transactions for higher stakes (Gambetta, 2009). Such institutional arrangements have an effect on how traders in a particular cryptomarket perceive the shadow of the future and, accordingly, on their strategies when interacting with each other (Axelrod, 1984; Guala, 2012).

On the one hand, these developments show that cryptomarket traders can never fully rely on platform providers to take the role of rule enforcers, making trust between buyers and sellers an essential part of maintaining cooperation in the Dark Web. On the other hand, such developments open avenues to analyse how institutional innovations affect cooperation between individuals in an otherwise largely de-centralized and unregulated environment. The ongoing growth and diversification of cryptomarkets (Soska and Christin, 2015), thus makes these Dark Web marketplaces particularly interesting for studying issues of cooperation and institutional evolution (Beckert and Wehinger, 2013).

Notes

1 In his seminal book, 'Order without Law: How Neighbors Settle Disputes', Ellickson (1991) describes how cooperation and social order are maintained in a close-knit community of cattle farmers in Shasta County, California. In his account, Ellickson repeatedly emphasizes the importance of gossip and reputation for promoting cooperative behaviour (e.g. p. 232): 'The residents of rural Shasta County gossip all the time. Indeed, any close-knit group is likely to have procedural norms that ask members to help spread truthful information about the prior prosocial or antisocial behavior of other members. By facilitating the flow of reputational information, these norms deter future uncooperative behavior by increasing an actor's estimates of the probability that informal enforcers would eventually catch up with him'. Although we study a 'community' of anonymous traders that engage in illegal market exchanges without ever meeting each other in person, the mechanisms by which cooperation is maintained are the same albeit digitalized in form of an electronic reputation system.

2 Note that this is different from the 'reputation' model put forward by Kreps *et al.* (1982). Kreps *et al.* (1982) start from the observation that in finitely repeated prisoners' dilemma (PD) games players cooperate more than the Nash equilibrium would predict. They explain this observation by suggesting that if players assume that with a small probability their interaction partner prefers to cooperate as long as they do, it may be rational for them to cooperate as well. In other words, a self-regarding player, who would defect in the one-shot PD, mimics a cooperative type in order not to forgo the higher benefits that result from several rounds of mutual cooperation as compared to mutual defection. Only once the sequence of interactions approaches the end, it becomes beneficial to defect and thereby reveal one's true type. Kreps and Wilson (1982) call this initial phase of mimicry 'reputation effect'. In contrast, we argue that reputation can be conceived as a costly signal because it is costly to acquire, which deters untrustworthy sellers to enter the market (Przepiorka and Berger, 2017).

3 Legal systems do not completely prevent scamming behaviour of online sellers, especially for transactions of small value, where potential costs of litigation become relatively high compared to potential losses for buyers. It has been estimated, based on buyer survey data, that 1–2 per cent of transactions in eBay contain fraudulent seller behaviour (Bauerly, 2009). While embeddedness of licit markets in legal systems cannot fully eradicate opportunistic behaviour, it provides strong assurances for buyers, which might account for a large part of trust facilitation that has been attributed to reputation systems in previous research.

4 Another way to address this question is to conduct laboratory experiments, in which a basic market environment can be staged, and the main tenets of the reputation mechanism can be put to an empirical test under controlled conditions (Falk and Heckman, 2009; Greiner and Ockenfels, 2009). The effectiveness of the reputation mechanism to promote trust and cooperation in social and economic exchange has been corroborated in a number of experiments (Bolton *et al.*, 2004; Yamagishi *et al.*, 2009; Kuwabara, 2015; Abraham *et al.*, 2016; although see Corten *et al.*, 2016). However, this approach still resembles the set-up of online markets for licit goods in many relevant respects. First, participants are recruited from a selective sample of mostly university students motivated to participate in experimental research. Secondly, although decisions in the laboratory are financially incentivized, the stakes are relatively low and potential losses limited to the opportunity costs of participation. Third, at the end of the day, participants are socially embedded in the environment of their universities which, in most cases, are embedded in well-functioning legal systems. In other words, actors are less likely to be blackmailed or even killed as a consequence of participating in experimental as compared to illegal markets.

Supplementary Data

Supplementary data are available at *ESR* online.

Acknowledgement

The authors are grateful to Nicolas Christin for providing them with the data and very helpful comments. They would also like to thank three anonymous reviewers for their perceptive comments and suggestions.

Funding

LN gratefully acknowledges support from the Netherlands Organisation for Scientific Research (NWO) [grant number: 406-12-004].

References

Abraham, M. *et al*. (2016). Reputation formation in economic transactions. *Journal of Economic Behavior & Organization*, **121**, 1–14.

Aldridge, J. and Décary-Hétu, D. (2016). Hidden wholesale: the drug diffusing capacity of online drug cryptomarkets. *International Journal of Drug Policy*, **35**, 7–15.

Allison, P. D. (2009). *Fixed Effects Regression Models*. Thousand Oaks, CA: SAGE.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books.

Bakos, Y. and Dellarocas, C. (2011). Cooperation without enforcement? A comparative analysis of litigation and online reputation as quality assurance mechanisms. *Management Science*, **57**, 1944–1962.

Barratt, M. J. and Aldridge, J. (2016). Everything you always wanted to know about drug cryptomarkets* (*but were afraid to ask). *International Journal of Drug Policy*, **35**, 1–6.

Bartlett, J. (2014). *The Dark Net: Inside the Digital Underworld*. London: Melville House.

Bauerly, R. J. (2009). Online auction fraud and eBay. *Marketing Management Journal*, **19**, 133–143.

Beckert, J. (2009). The social order of markets. *Theory and Society*, **38**, 245–269.

Beckert, J. and Wehinger, F. (2013). In the shadow: illegal markets and economic sociology. *Socio-Economic Review*, **11**, 5–30.

Bolton, G. E., Greiner, B. and Ockenfels, A. (2013). Engineering trust: reciprocity in the production of reputation information. *Management Science*, **59**, 265–285.

Bolton, G. E., Katok, E. and Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*, **50**, 1587–1602.

Bowles, S. and Gintis, H. (2011). *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton: Princeton University Press.

Bozoyan, C. and Vogt, S. (2016). The impact of third-party information on trust: valence, source, and reliability. *PLoS One*, **11**, e0149542.

Branwen, G. (2016). Black-market risks: analysis of lifespans, deaths, and predictive factors, *gwern.net* Blogs. Available from: <http://www.gwern.net/Black-market%20survival> [accessed 19 December 2016].

Calkins, M. M., Nikitkov, A. and Richardson, V. (2008). Mineshafts on Treasure Island: a relief map of the eBay fraud landscape. *Pittsburgh Journal of Technology Law and Policy*, **8**, 1–52.

Carruthers, B. G. (2013). From uncertainty toward risk: the case of credit ratings. *Socio-Economic Review*, **11**, 525–551.

Christin, N. (2012). Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. *arXiv*, 1207.7139.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: The Belknap Press of Harvard University Press.

Corten, R. *et al*. (2016). Reputation effects in social networks do not promote cooperation: an experimental test of the Raub & Weesie model. *PLoS One*, **11**, e0155703.

Dasgupta, P. (1988). Trust as a commodity. In Gambetta, D. (Ed.), *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell, pp. 49–72.

Décary-Hétu, D., Paquet-Clouston, M. and Aldridge, J. (2016). Going international? Risk taking by cryptomarket drug vendors. *International Journal of Drug Policy*, **35**, 69–76.

Dellarocas, C. (2003). The digitization of word of mouth: promise and challenges of online feedback mechanisms. *Management Science*, **49**, 1407–1424.

Diekmann, A. *et al*. (2014). Reputation formation and the evolution of cooperation in anonymous online markets. *American Sociological Review*, **79**, 65–85.

Diekmann, A. and Przepiorka, W. (forthcoming). Trust and reputation in markets. In Giardini, F. and Wittek, R. (Eds.), *The Oxford Handbook of Gossip and Reputation*. Oxford: Oxford University Press.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, **8**, 100–110.

Ellickson, R. C. (1991). *Order without Law: How Neighbors Settle Disputes*. Cambridge: Harvard University Press.

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, **326**, 535–538.

Feinberg, M., Willer, R. and Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science*, **25**, 656–664.

Fligstein, N. (2001). *The Architecture of Markets: An Economic Sociology of Twenty-First-Century Capitalist Societies*. Princeton, NJ: Princeton University Press.

Frey, V. and van de Rijt, A. (2016). Arbitrary inequality in reputation systems. *Scientific Reports*, **6**, 38304.

Friedman, E. J. and Resnick, P. (2001). The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy*, **10**, 173–199.

Gambetta, D. (2009). *Codes of the Underworld: How Criminals Communicate*. Princeton: Princeton University Press.

Global Drug Survey. (2016). *Key findings from the Global Drug Survey 2016*. London: Global Drug Survey. Available from: <https://www.globaldrugsurvey.com/past-findings/the-global-drug-survey-2016-findings/> [accessed 12 December 2016].

Granovetter, M. (1992). Problems of explanation in economic sociology. In Nohria, N. and Eccles, R. G. (Eds.), *Networks and Organizations: Structure, Form, and Action*. Boston: Harvard Business School Press, pp. 25–56.

Greif, A. (1989). Reputation and coalitions in medieval trade: evidence on the Maghribi traders. *Journal of Economic History*, **49**, 857–882.

Greiner, B. and Ockenfels, A. (2009). Vom Labor ins Feld: Die Ökonomik des Vertrauens. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie, Sonderheft*, **49**, 219–242.

Grimalda, G., Pondorfer, A. and Tracer, D. P. (2016). Social image concerns promote cooperation more than altruistic punishment. *Nature Communications*, **7**, 12288.

Guala, F. (2012). Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, **35**, 1–15.

Güth, W. and Ockenfels, A. (2003). The coevolution of trust and institutions in anonymous and non-anonymous communities. In Holler, M.J., Kliemt, H., Schmidtchen, D. and Streit, M. (Eds.), *Jahrbuch für Neue Politische Ökonomie*. Tübingen: Mohr Siebeck, pp. 157–174.

Hardy, R. A. and Norgaard, J. R. (2016). Reputation in the internet black market: an empirical and theoretical analysis of the deep web. *Journal of Institutional Economics*, **12**, 515–539.

Hillmann, H. (2013). Economic institutions and the state: insights from economic history. *Annual Review of Sociology*, **39**, 251–273.

Hillmann, H. and Aven, B. L. (2011). Fragmented networks and entrepreneurship in late imperial Russia. *American Journal of Sociology*, **117**, 484–538.

Jappelli, T. and Pagano, M. (2002). Information sharing, lending and defaults: cross-country evidence. *Journal of Banking and Finance*, **26**, 2017–2045.

Kollock, P. (1994). The emergence of exchange structures: an experimental study of uncertainty, commitment, and trust. *American Journal of Sociology*, **100**, 313–345.

Kollock, P. (1999). The production of trust in online markets. In Lawler, E. J. and Macy, M. W. (Eds.), *Advances in Group Processes*, Vol. **16**. Amsterdam: JAI Press, pp. 99–123.

Kreps, D. M. *et al*. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, **27**, 245–252.

Kreps, D. M. and Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, **27**, 253–279.

Kuwabara, K. (2015). Do reputation systems undermine trust? Divergent effects of enforcement type on generalized trust and trustworthiness. *American Journal of Sociology*, **120**, 1390–1428.

Lindenberg, S. (2017). The dependence of human cognitive and motivational processes on institutional systems. In Jann, B. and Przepiorka, W. (Eds.), *Social Dilemmas, Institutions and the Evolution of Cooperation*. Berlin: De Gruyter Oldenbourg, pp. 85–106.

Martin, J. (2014). *Drugs on the Dark Net: How Cryptomarkets are Transforming the Global Trade in Illicit Drugs*. Basingstoke: Palgrave Macmillan.

Matzat, U. and Snijders, C. (2012). Rebuilding trust in online shops on consumer review sites: sellers' responses to user-generated complaints. *Journal of Computer-Mediated Communication*, **18**, 62–79.

Milgrom, P. R., North, D. C. and Weingast, B. R. (1990). The role of institutions in the revival of trade: the law merchant, private judges, and the champagne fairs. *Economics and Politics*, **2**, 1–23.

Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philosophical Transactions of the Royal Society B*, **371**, 20150100.

Milinski, M., Semmann, D. and Krambeck, H.-J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature*, **415**, 424–426.

Nee, V. (2005). The new institutionalisms in economics and sociology. In Smelser, N. J. and Swedberg, R. (Eds.), *The Handbook of Economic Sociology*. Princeton, NJ: Princeton University Press, pp. 49–74.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, **314**, 1560–1563.

Pavlou, P. A. and Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, **15**, 37–59.

Przepiorka, W. (2013). Buyers pay for and sellers invest in a good reputation: more evidence from eBay. *Journal of Socio-Economics*, **42**, 31–42.

Przepiorka, W. and Aksoy, O. (2017). *Social Order in Online Markets and the 'Collapse' of Institutions*. Unpublished manuscript, The Netherlands: Department of Sociology/ICS, Utrecht University.

Przepiorka, W. and Berger, J. (2017). Signalling theory evolving: signals and signs of trustworthiness in social exchange. In Jann, B. and Przepiorka, W. (Eds.), *Social Dilemmas, Institutions and the Evolution of Cooperation*. Berlin: De Gruyter Oldenbourg, pp. 373–392.

Resnick, P. and Zeckhauser, R. (2002). Trust among strangers in internet transactions: empirical analysis of eBay's reputation system. In Baye, M. R. (Ed.), *The Economics of the Internet and E-Commerce*. Amsterdam: Elsevier, pp. 127–157.

Shapiro, C. (1983). Premiums for high quality products as return to reputation. *Quarterly Journal of Economics*, **98**, 659–680.

Snijders, C. and Weesie, J. (2009). Online programming markets. In Cook, K. S., Snijders, C., Buskens, V. and Cheshire, C. (Eds.), *eTrust: Forming Relationships in the Online World*. New York, NY: Russell Sage Foundation, pp. 166–185.

Sommerfeld, R. D. *et al*. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences of the USA*, **104**, 17435–17440.

Soska, K. and Christin, N. (2015). Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In 24th USENIX Security Symposium (USENIX Security 15). Washington, DC: USENIX Association, pp. 33–48.

Standifird, S. S. (2001). Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings. *Journal of Management*, **27**, 279–295.

Sylwester, K. and Roberts, G. (2013). Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evolution and Human Behavior*, **34**, 201–206.

Willer, R. (2009). Groups reward individual sacrifice: the status solution to the collective action problem. *American Sociological Review*, **74**, 23–43.

Wu, J., Balliet, D. and van Lange, P. A. M. (2016). Gossip versus punishment: the efficiency of reputation to promote and maintain cooperation. *Scientific Reports*, **6**, 23919.

Yamagishi, T. *et al*. (2009). Solving the lemons problem with reputation. In Cook, K. S., Snijders, C., Buskens, V. and Cheshire, C. (Eds.), *eTrust: Forming Relationships in the Online World*. New York, NY: Russell Sage Foundation, pp. 73–108.

Wojtek Przepiorka is Assistant Professor at the Department of Sociology in Utrecht. Before moving to the Netherlands, he held a research fellow position at Nuffield College and the Department of Sociology in Oxford. His research interests are in analytical and economic sociology, game theory, and quantitative methodology, in particular experimental methods. His most recent publications include '"Take One for the Team!" Individual heterogeneity and the emergence of latent norms in a volunteer's dilemma' (*Social Forces* 94, with A. Diekmann), and 'Generosity is a sign of trustworthiness—the punishment of selfishness is not' (*Evolution and Human Behavior*, 37, with U. Liebe).

Lukas Norbutas is a PhD candidate at the Department of Sociology of Utrecht University, and the Netherlands Institute for the Study of Crime and Law Enforcement. His research interests include economic sociology, social network analysis, cooperation and trust, and online criminal networks.

Rense Corten is Associate Professor at the Department of Sociology of Utrecht University. His research revolves around the themes of cooperation, trust, and (the dynamics of) social networks, with empirical applications including adolescent networks, social media, the sharing economy, online criminal networks, and laboratory experiments. His work has been published in journals such as the *American Sociological Review*, *Social Networks*, *Rationality and Society*, and others.