9-2010

DOES REPUTATION REALLY SIGNAL POTENTIAL SUCCESS IN ONLINE MARKETPLACES, OR IS IT ONLY A TRIGGER?

David Gefen
Drexel University, USA, gefend@drexel.edu

Erran Carmel
American University, USA, carmel@american.edu

Does Reputation Really Signal Potential Success in Online Marketplaces, or is it only a Trigger?

David Gefen, Drexel University and Erran Carmel American University

Abstract

The implicit assumption in online marketplaces is that reputation, registered as the average of previous ratings, represents a market opinion of the trustworthiness of a business party and that this predicts eventual satisfaction with the expected outcome of the transaction. Extensive research indeed shows that such reputation does result in a higher probability of the seller being chosen in both services and goods markets, presumably because buyers believe that higher rating sellers have a better probability of delivering. Whether reputation actually predicts successful completion of the project and payment, rather than bid choice, however, is an unknown. This study answers that question. The data show that only when the rating value is above 5.5, which is rounded up in the market presentation to a caption of "above average", is there a significant relationship between previous rating of the seller and eventual payment.

Keywords: Trust, Contracting, Service Auctions, software online markets, Buyer Feedback, Seller Ratings, Buyer Ratings.

1 ON THE IMPORTANCE OF TRUSTWORTHINESS INDICATORS IN ONLINE MARKETS

The trust a buyer has in an individual seller in a marketplace, which in the case of online software markets means the coders who are paid to develop the software, is crucial in determining which bid is chosen (Gefen & Carmel, 2008). Trust is the willingness to depend upon the actions of another person in situations involving vulnerability and dependency (Mayer, Davis, & Schoorman, 1995), which is precisely the case with software development (Gefen, 2002, 2004; Gefen, Wyss, & Lichtenstein, 2008). Trust is the assumption that the trusted party will not behave opportunistically by taking advantage of the situation (Gefen, Karahanna, & Straub, 2003), and will fulfill its commitments (Luhmann, 1979). Trust determines, in part, behavioral intentions to engage in an activity with the trusted party in situations characterized by there being an element of risk concerning the behavior of this trusted party. This has been widely shown in ecommerce and online product auctions (Choudhury & Karahanna, 2008; Gefen, et al., 2003; Kim, Ferrin, & Rao, 2009). People trust because they think the trusted party is trustworthy; Being trustworthy means this person showed integrity, ability, and benevolence (Mayer, et al., 1995).

Applying this definition to online software markets, a trustworthy coder is one who the buyer believes is able to deliver, showing ability, and, when applicable, demonstrate appropriate caring about the buyer, showing benevolence, while keeping promises, showing integrity. Trusting this coder means the buyer is willing to take the risk of depending on the coder, i.e. giving this coder the tender (Gefen & Carmel, 2008).

To a large extent this perception of trustworthiness is based on past experience of both of the trusting person (Gefen, 2000; Luhmann, 1979) and of others (Dellarocas, 2003). Indeed, trust in online markets is often built based at least party through the previous ratings made by other buyers (Dellarocas, 2003), as also shown in the content of comments attached to these ratings (Pavlou & Dimoka, 2006). Ratings as a trustworthiness cue also command a premium (Ba & Pavlou, 2002; Pavlou & Dimoka, 2006) and are considered by some as an equivalent of word of mouth recommendation (Dellarocas, 2003).

Research to date has not shown however whether ratings actually results in the transaction ultimately being paid, which is the real dependent variable of interest in services markets.

Allowing that ratings represent market indicators of trustworthiness, an argument can be made why rating should behave as a trigger. Behaving as a trigger means that only ratings above a certain critical value should increase the identification of eventual payment, i.e. delivery of the acquired software service to the satisfaction of the buyer. The reason for this is that trust only comes into effect when the level of perceived risk can be overcome by trusting. When risk is too high relative to the level of trust, such as when ratings are low, trust should be immaterial (Mayer, et al., 1995). And so, presumably, among those low rating coders who are nonetheless chosen by the buyers, trust based on previous trustworthiness was not the primary reason they were chosen and should not, therefore, predict success the way trust usually does. Previous research on trust has not shown such a trigger effect.

2 DATA COLLECTION

The data came from RentACoder.com. This marketplace gave us a copy of their archival data for all transactions up to and including 2005.[1] The data analyzed here contain all the tenders and all the bids made in 2005. Previous average and number of coder rating, previous average and number of buyer ratings, and the number of previous contracts between the buyer and the seller were all generated back to when the website began. In 2005 there were 320,790 bids made by 4544 sellers on 14,433 tenders posted

[1] Gefen and Carmel (2008) used an older data set.

by 5,874 coders. On average these seller made 3.18 bids (std. 4.97) bids and the buyers posted 2.46 tenders (std.=3.80) in 2005. 11,794 of these tender, or 81.7%, were paid. There were between 1 and 274 bids on each tender, mean 26.28 with std. 25.41. Ratings ranged from 1 low to 10 highest. The average bid amount was $149.31 (std. 405.15).

3 DATA ANALYSIS

Analysis of the data supports the trigger proposition and thus qualifies the implied assumption in the marketplace that ratings predict payment. Only tenders below $400 were included and only those with at least 2 bids on them. Examining whether the average of seller ratings predict whether the tender is eventually paid, the data show that higher average seller ratings, but not buyer ratings, contribute strongly to identifying which projects will be paid and which not, but that the story is more complex. Only when seller rating average was in or above in the market defined “above average” category did seller average ratings contribute to identifying which projects will be paid. And, even then seller ratings contributed mainly to identifying which tenders were eventually paid, and contributed almost nothing to identifying which tenders were not eventually paid.

Examining the comments associated with these seller ratings shows the predominance of keywords dealing with ability and dependability, indicating the importance of trustworthiness in this process, supporting the theory base used to predict this trigger effect.

The data also indicate that the bidding characteristics controls included in the study, including higher priced tenders, bidding above the average of the tender, and more sellers bidding all contribute mainly to increase the identification of unpaid tenders but add up to only an R^2 equivalent of 9%. Adding the count and average of seller rating next bounces the equivalent to over 21%, identifying mainly the paid tenders.

The key finding of the trigger effect is shown in Figure 1. The figure shows the percent of paid tenders for each of the grouped values of previous seller ratings. The colors indicate marketplace categorization of rating numbers, shown in Figure 2. These numbers range from 1 to 10, with 6 being “above average”. As the figure suggests, and logistic regressions in Table 1 show, only when the ranking is above 5.5, which is rounded up in the market presentation to “above average”, is there a significant relationship between rating and the percent of paid transactions at that rating level.

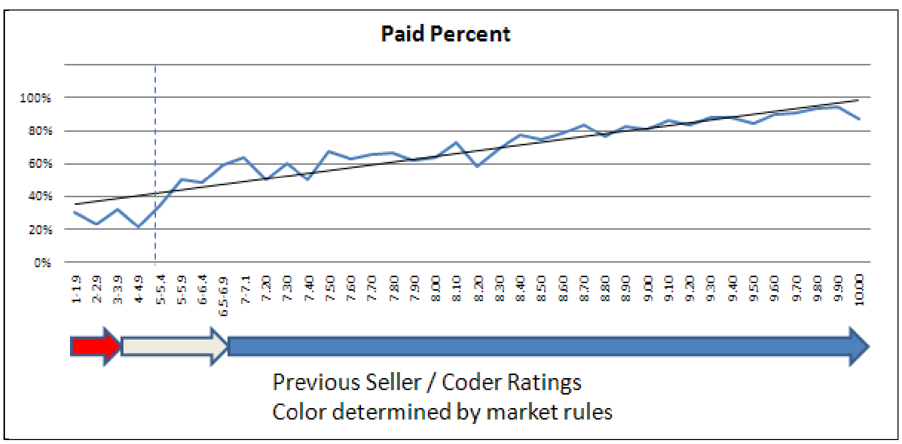
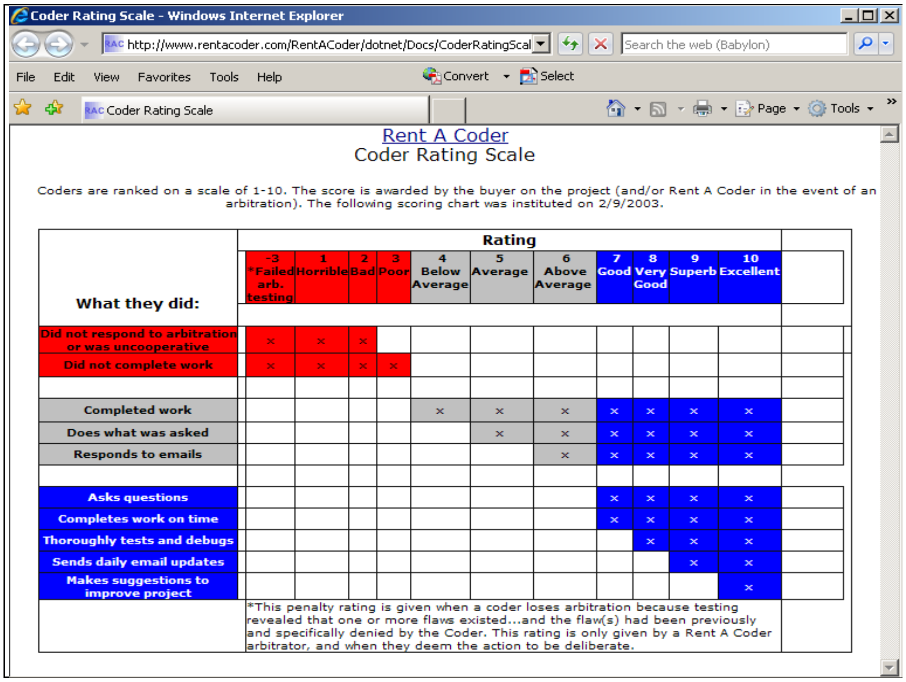


Figure 1. Percent of Paid Transactions by Seller Rating



Coder Rating Scale - Windows Internet Explorer

http://www.rentacoder.com/RentACoder/dotnet/Docs/CoderRatingScal

Rent A Coder

Coder Rating Scale

Coders are ranked on a scale of 1-10. The score is awarded by the buyer on the project (and/or Rent A Coder in the event of an arbitration). The following scoring chart was instituted on 2/9/2003.

What they did:	-3 *Failed arb. testing	1 Horrible	2 Bad	3 Poor	4 Below Average	5 Average	6 Above Average	7 Good	8 Very Good	9 Superb	10 Excellent
Did not respond to arbitration or was uncooperative	×	×	×								
Did not complete work	×	×	×	×							
Completed work					×	×	×	×	×	×	×
Does what was asked						×	×	×	×	×	×
Responds to emails							×	×	×	×	×
Asks questions								×	×	×	×
Completes work on time								×	×	×	×
Thoroughly tests and debugs									×	×	×
Sends daily email updates										×	×
Makes suggestions to improve project											×

*This penalty rating is given when a coder loses arbitration because testing revealed that one or more flaws existed...and the flaw(s) had been previously and specifically denied by the Coder. This rating is only given by a Rent A Coder arbitrator, and when they deem the action to be deliberate.

Figure 2. *Seller Rating Values and Color Coding*

The data were analyzed with logistic regression. The dependent variable was whether a payment was made on this tender. The controls were entered in block 1 and seller ratings and the ln of their count in the second block. Sensitivity analysis was also done. Checking previous average seller rating at various values below 5.5, such as 5.4, 5.3, 5.2, and so on, produced the same pattern as when the average is set to below 5.5 in column 2. Likewise, setting the critical value of previous average seller rating above 5.5, such as 5.6, 5.7, sand 6.0, produced the same pattern as when the average is set to above 5.5 in column 3. The average seller ratings were examined in increments of .1 between 4 and 7. Table 1 shows these results together with analysis of all the data in column 1.

4 CONTRIBUTION

The study contributes to theory and practice by: (1) showing that rating is a trustworthiness indicator, and (2) that it behaves in a trigger manner, with only above average values contributing to identifying successful transactions, but (3) that even so it mainly identifies the paid, but much less so the more important category of unpaid transactions. (4) This important category of unpaid transactions is identified by the bidding characteristics. Moreover, (5) lowering bid prices through bidding may actually be counterproductive in that lower priced bids and bids below the market average are more likely to end up unpaid. (6) Additionally, the data show that count and average buyer ratings are an insignificant predictor of whether the transaction will be paid or not.

Table 2. *Results of Logistic Regression. Only Independent Variables included by the Stepwise Procedure are Shown.*

	All the data	Previous Seller Rating Average <=5.4	Previous Seller Rating Average >5.5
	Column 1	**Column 2**	**Column 3**
Sample Size	12030	327	11662
of these unpaid	1863	227	1615
Percent Paid	85%	31%	86%
Block 1 Wald Stepwise			
# previous contracts	1.99 (p<.001)		1.83 (p<.001)
ln Amount	-.31 (p<.001)	-.47 (p<.001)	-.32 (p<.001)
Amount Ratio to Average	.62 (p<.001)		.59 (p<.001)
Days to completion	-.01 (p<.001)		-.004 (P<.001)
How many coders bidding	.03 (p<.001)		.03 (p<.001)
Average Buyer ratings			
Constant	2.19 (p<.001)	1.04 (p=.014)	2.41 (p<.001)
Nagelkerke R^2	0.092	0.137	0.089
Correct Classification Unpaid	71.6	57.3	68
Correct Classification Paid	**54.3**	64	59.1
Correct Classification Total	57	59.3	60.3
Block 2 Wald Stepwise			
# previous contracts	1.75 (p<.001)		1.70 (p<.001)
ln Amount	-.36 (p<.001)	-.47 (p<.001)	-.35 (p<.001)
Amount Ratio to Average	.42 (p<.001)		.39 (p<.001)
Days to completion	-.00 (p<.001)		-.004 (P<.001)
How many coders bidding	.02 (p<.001)		.02 (p<.001)
Average Buyer ratings	not included		
Ln Number of previous coder ratings	*.24 (p<.001)*	-.55 (P=.005)	.27 (p<.001)
Average of previous coder ratings	*.41 (p<.001)*		.59 (p<.001)
Constant	-1.61 (p<.001)	1.30 (p=.003)	3.42 (p<.001)
Nagelkerke R^2	0.211	0.172	0.161
Correct Classification Unpaid	68.9	66.1	67.8
Correct Classification Paid	**70.8**	60	69.3
Correct Classification Total	70.5	64.2	69.1

5 REFERENCES

Ba, S., & Pavlou, P. A. (2002). Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly*, *26*(3), 243-268.

Choudhury, V., & Karahanna, E. (2008). The Relative Advantage of Electronic Channels: A Multidimensional View. *MIS Quarterly, 32*(1), 179-200.

Dellarocas, C. (2003). The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, *49*(10), 1407-1424.

Gefen, D. (2000). E-commerce: The Role of Familiarity and Trust. *Omega*, *28*(6), 725-737.

Gefen, D. (2002). Nurturing Clients' Trust to Encourage Engagement Success During the Customization of ERP Systems. *Omega: The International Journal of Management Science, 30*(4), 287-299.

Gefen, D. (2004). What Makes ERP Implementation Relationships Worthwhile: Linking Trust Mechanisms and ERP Usefulness. *Journal of Management Information Systems, 23*(1), 263-288.

Gefen, D., & Carmel, E. (2008). Is the World Really Flat? A Look at Offshoring in an Online Programming Marketplace. *MIS Quarterly*, *32*(2), 367-384.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly*, *27*(1), 51-90.

Gefen, D., Wyss, S., & Lichtenstein, Y. (2008). Business Familiarity as Risk Mitigation in Software Development Outsourcing Contracts. *MIS Quarterly*, *32*(3), 531-551.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2009). Trust and Satisfaction, Two Stepping Stones for Successful E-Commerce Relationships: A Longitudinal Exploration. *Information Systems Research*, *20*(2), 237-257.

Luhmann, N. (1979). *Trust and Power* (T. f. German, Trans.). Great-Britain: John Wiley and Sons.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709-734.

Pavlou, P. A., & Dimoka, A. (2006). The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Information Systems Research*, *17*(4), 392-414.