Spiral of Hatred: Social Effects in Buyer-Seller Cross-Comments Left on Internet Auctions*

Radoslaw Nielek[1], Aleksander Wawer[2], and Adam Wierzbicki[3,**]

[1] Polish Japanese Institute of Information Technology,
ul. Koszykowa 86, 02-008 Warszawa, Poland
nielek@pjwstk.edu.pl

[2] Institute of Computer Science Polish Academy of Sciences,
ul. J.K. Ordona 21, 01-237 Warszawa, Poland
axw@ipipan.waw.pl

[3] Polish Japanese Institute of Information Technology,
ul. Koszykowa 86, 02-008 Warszawa, Poland
adamw@pjwstk.edu.pl

Abstract. An auction platform is a dynamic environment where a rich variety of social effects can be observed. Most of those effects remain unnoticed or even hidden to ordinary users. The in-depth studies of such effects should allow us to identify and understand the key factors influencing users' behaviour. The material collected from the biggest Polish auction house has been analyzed. NLP algorithms were applied to extract sentiment-related content from collected comments. Emotional distance between negative, neutral and positive comments has been calculated. The obtained results confirm the existence of the spiral-of-hatred effect but also indicate that much more complex patterns of mutual relations between sellers and buyers exist. The last section contains a several suggestions which can prove useful to improve trustworthiness of users' reports and security of an auction platform in general.

1 Introduction

Transaction volumes and numbers of users in e-commerce systems have been booming over the past few years and there is no sign of a slowdown in the foreseeable future. Every new account in an auction house or within web 2.0 services creates new challenges for privacy and security. Auction houses seem to be the most demanding environment for trust management systems due to direct relationship between reputation and users' income [3][4]. Every unpunished and undetected fraud undermines the honest agents' motivation to play fair. Thus, many researchers are working to create new reputation algorithms. Nevertheless, reputation management systems embedded in the most popular websites remain practically unchanged over years and are based on very simple quantitative evaluations and qualitative comments.

* This research has been supported by the Polish Ministry of Science grant N N516 4307 33.

** This author has been supported by the grant of the Polish Ministry of Science 69/N-SINGAPUR/2007/0.

E. Ferrari et al. (Eds.): TM 2009, IFIP AICT 300, pp. 1–14, 2009.

Thus far, most researches have been focused on improving algorithms using qualitative feedback and therefore there is a relatively narrow selection of papers devoted to mining comments (security perspective [2], trustworthiness of reviews [17]) and developing algorithms for trust management systems which explicitly consider descriptive opinions [15]. This is so partly because natural language processing module, which is the cornerstone of such an algorithm, requires building it almost from scratch for every single language (the reusable part is insignificant). It means that the results obtained for languages other than English are hardly comparable and difficult to validate for a larger scientific community.

Nevertheless, it makes good sense to devote resources to the discovery of patterns in descriptive opinions expressed in languages other than English since most Internet transactions are done in language environment native to participants and as local web auction markets grow very fast, this situation will probably continue into the future. Many observations reported in this paper are likely to apply to other cultures too, irrespective of the language in which the comments are written. Primarily, users' behavioural patterns refer to more general psychological (e.g. spiral of hatred - response is stronger than impulse [14]) and sociological effects which can be even stronger than the cultural fingerprint. A comparative study on Taobao (Chinese version of an online auction marketplace) and eBay has partly confirmed this assumption [10].

In reference to the above, this paper is devoted to an analysis of users' behaviour during the after-transaction evaluation process, in particular taking into account pairs of comments on the same transactions delivered by both sellers and buyers (cross-comments) in the auction house. Two approaches have been used to identify and validate different hypotheses. In section 2, feedback mechanisms existing in e-commerce systems are described. Section 3 is devoted to quantitative and statistical examination of the collected data and focuses on the effects related to comment type and order in which they arrive. The results obtained by natural language processing algorithms in the context of the hypothesis for validating the spiral of hatred effect are presented in the fourth section. The fifth section features discussion of the results and a new heuristic model to solve some of the identified problems. The last section presents the conclusions and possible trends in the future research.

2 Quantitative and Qualitative Comments

The most commonly used reputation systems embedded in online auction website allow us to evaluate transaction results not only by selecting a predefined category from a list but also by leaving shorter or longer comments. The quantitative measurement in use by eBay and Allegro (the biggest Polish auction house) is based on a very simple structure. When a given transaction is completed, every eBay/Allegro user can evaluate his or her partner by choosing either a positive or neutral, or negative mark. The evaluation mark is visible after being submitted. On eBay it is also possible to evaluate separately the quality of a delivered product, communication, shopping time as well as shipping and handling charges. All those additional evaluation are anonymous. The sums of positive, negative and neutral marks are presented separately. Because feedback is not obligatory, not every transaction is followed by its evaluation. As shown in [1] no information is usually indicative of bad experience during the transaction.

Predominantly, only positive comments appear. For more than 1.7 million comments in the collected database there were only ca. 9000 negative and ca. 5000 neutral comments which means that either the fraud level is very low or (it seems more likely) there is a mechanism, which discourages people from making negative comments. Certainly, the threat of legal action [8] constitutes one source of fear, another one is probably related to the possibility of being punished with negative reciprocal evaluation. Yet, another effect identified by researchers [7] is that users award a positive quantitative evaluation mark but describe all negative aspects of a transaction in words.

The relative stable framework in the auction houses provides a good opportunity to detect even quite complicated users' behavioural patterns. Abilities of users to learn from previous experiences and to modify their strategies appear to be non-trivial attractors within the space of possible behavioural patterns. A good example of self-adaptation in the complex system which has emerged in online auction websites is that users pay much more attention to negative comments when they calculate transaction risk [11].

Typically, users can intentionally express their opinions only by making comments which are composed of a selected label (quantitative) and a description (qualitative). Nevertheless, a lot of additional information can be found in the data collected in the online auction website, for example response times on positive and negative evaluations, order of buyer-seller evaluation, length of comments or context and reference points (average rating for specific subsets). Identifying measurable effects in buyer-seller interaction can help improve the existing trust management algorithms and create a foundation for designing new ones.

3 Experiments

3.1 Dataset

The database analysed in this article was provided by the biggest Polish auction house (over 70% of market share). At the beginning of the fourth quarter in 2006, 10,000 sellers and 10,000 buyers have been randomly selected; their profiles and received comments have been stored (description and evaluation). During the next 6 months all transactions conducted by the selected users were monitored and recorded. For every partner who appeared in transaction and was not in the primary database, all historical information about the received feedback has been collected, but with respect to new auctions only the originally selected users have been monitored. In the first quarter in 2007 the database contained more than 200,000 transactions and over 1.7 million comments.

3.1.1 Formal Definition

Symbols used in the following sections are defined below:

- U — set of all users,
- T — set of all transaction,
- t_m — m-th transaction,
- u_i — i-th user,
- $c_{t_m}^{u_i}$ — comment left by the i-th user after m-th transaction,

- $\tau(c_{tm}^{u_i})$ — sentiment measured by sentipejd for the comment c,
- $\rho(c_{tm}^{u_i})$ — label for the comment c given by the ith user,
- $r(t_m, u_i)$ — the role of the i-th user in the m-th transaction (either buyer or seller),
- $\varphi(c_{tm}^{u_i})$ — timestamp for the comment c,
- $\omega_{t_m} = (c_{tm}^{u_i}, c_{tm}^{u_j})$ — an ordered pair of comments for the m-th transaction,
- $\delta(\omega_{t_m})$ — time between two comments,

3.1.2 Amount, Type, Time and Order in Cross-Comments

The objective of this paper is to identify the effects which appear during bi-directional evaluation, therefore the main focus was an analysis of the ordered pairs of comments, defined in the previous section as ωt_m. For over 1.7 million comments slightly more than 800 thousand pairs were found (in ca. 9% of cases only one party of a given transaction left a comment – either buyer or seller) Only 5056 of pairs contain at least one non-positive evaluation.

Over 90% of answers for comments are made within 14 days after the first evaluation. Shape of curves on the Fig.1 is similar for all considered cases but there is a notable bias in the starting point. In general, sellers are more responsive - for negative and neutral comments over 20% of sellers and only 7% of buyers feedbacks were written in less than one hour after receiving an evaluation from the partner (for positive comments the numbers are 7% and 3% respectively). On average, buyers seem to visit the auction website less often, so their reaction is slower. Comments, regardless of their contents, are emailed to the evaluated user, thus there is no other variable, except for the type of comment, that may explain the variation in reaction times. Very short response times for negative and neutral comments (when compared with positive

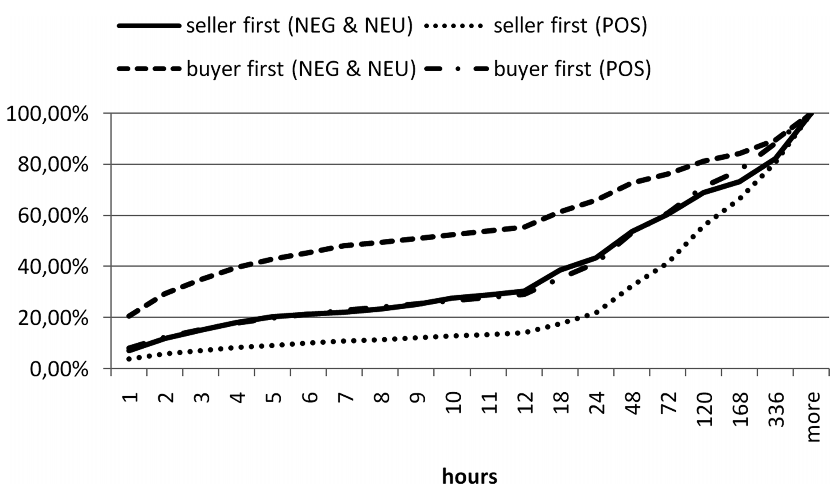

Fig. 1. Time-span between comment and answer for different category of evaluation (cumulative histogram)

Table 1. Average length of comments (in characters)

	Buyer	Seller
POS	102.49	73.18
NEU	149.02	154.27
NEG	183.77	178.78

feedbacks) can be explained by the will to punish, as fast as possible, the author of the negative[1] evaluation.

Average length of comments presented in Table 1. varies between both transaction roles and different feedback types. As a rule, longer text creates an opportunity to enumerate more facts and express a broader variety of emotions, but also emphasizes the importance of the particular comment for a given user - she or he has been ready to devote more time to leave feedback. For a positive experience, which is a typically expected result of the transaction, the comments are relatively short - 100 characters in the case of sellers and 73 of buyers (the difference is statistically significant). Manual inspection of a both comment types indicates that the cause of this difference results from the habit of adding an advertisement at the end of comments made by sellers (e.g. *"Hope to see you again. ALFRA_PL"*).

Dissatisfying transaction outcome is positively correlated with length of evaluations. More characters are needed to describe and probably justify dissatisfaction and the negative feedback. The difference between buyer and seller observed for positive comments disappears for both negative and neutral evaluations (small differences observed in table 1. in the second and third row are statistically insignificant).

The unwritten rule in online auction websites is that buyers make comments first. For the pairs of comments containing only positive evaluation in 8.2% cases this rule was broken. If one of the comments is negative or neutral, the number of cases contradicting the unwritten law rises dramatically to over 18%. There are many reasons why sellers decide to make a comment first. Some of the sellers probably participate in too many transactions to follow which one is already finished and commented and which not. Strong evidence of such behaviour can be seen in Table 2 in the very last column - more than 8% of sellers answered to negative evaluation with a positive one. More detailed manual analysis of these pairs indicates that some of the answers contain explanations of the reason for unsatisfactory quality of service (e.g. limited access to the Internet or problems with logistic) but most is given disregarding the previous

Table 2. Combination of comment-answer pairs (buyer first)

	POS/x	NEU/x	NEG/x
x/POS	—	937 (22.60%)	339 (8.17%)
x/NEU	0 (0%)	686 (16.55%)	41 (0.99%)
x/NEG	0 (0%)	408 (9.84%)	1734 (41.83%)

[1] As it is shown in the next sections, neutral comments are very similar to negative comments.

negative comment. Yet another hypothesis is that the seller is forced by an external event to send the feedback- he or she needs to pay commission for the auction website within a limited period of time after finishing the transaction regardless of its outcome.

If the buyer is satisfied and expresses this satisfaction with a positive comment first, the answer from the seller will always be positive. All the collected pairs confirm this rule without exceptions. It could be only partially explained by the previous observation. First, a vast domination of positive comments makes a pair $\omega_{t_m} = (c^{u_i}_{t_m}, c^{u_j}_{t_m})$ such as $\rho(c^{u_i}_{t_m}) = pos \wedge \rho(c^{u_i}_{t_m}) \neq pos$ statistically very improbable. Yet, the distribution of such comments was asymmetrical between both transaction roles. On one hand, of over such 550 cases exist for $r(t_m, u_i) = seller$, on the other, no such comment pair was found for $r(t_m, u_i) = buyer$. Secondly, the results[11] show that even substantial amount of negative feedback does not affect the ability of buyers to participate in transactions. Therefore, a positive opinion about the seller is always rewarded with a reciprocal positive feedback. Thirdly, although there is no explicitly defined procedure to change already submitted feedback, it is essentially possible after some reasonable efforts (e.g. sending an email to the webmaster). So, a seller can refrain from making a non-positive evaluation only because of an aversion to initiating a "war", even though not everything went correct during the transaction.

Table 3. Combination of comment-answer pairs (seller first)

	POS/x	NEU/x	NEG/x
x/POS	—	0 (0%)	0 (0%)
x/NEU	284 (31.17%)	19 (2.08%)	19 (2.08%)
x/NEG	242 (26.56%)	8 (0.87%)	339 (37.21%)

Ordered pairs of comments $\omega_{t_m} = (c^{u_i}_{t_m}, c^{u_j}_{t_m})$ such as $r(t_m, u_i) = neu \wedge r(t_m, u_j) = neg$ (the first evaluation is neutral and the second negative) appear eight times more frequently (416) than pairs where $r(t_m, u_i) = neg \wedge r(t_m, u_j) = neu$ (the first evaluation is negative and the second neutral) (60). This enormous disproportion cannot be explained by the course of the transaction because there is no evidence to claim that a negatively affected party will comment second. A more credible explanation is that neutrally evaluated agents use negative evaluations as a punishment and try to do it as severely as possible.

4 Mining the Meaning of Comments

4.1 Automatic Sentiment Extraction

For the sentiment analysis task we used a modified version of Sentipejd [20] - a hybrid of lexeme category analysis with a shallow parsing engine. At the basic level, Sentipejd checks for presence of a specific category of lexemes. Such an abstraction originates in content analysis systems, most notably the classic General Inquirer [21]. Lexical categories used in this work include two sets of words (dictionaries): 1580 positive and

1870 negative ones, created by Zetema[2]. Because comment texts are typed in a careless manner, very often completely without diactrits, lexeme recognition was extended with a diacrit guesser. Recognized sentiment lexemes, along with morphosyntactic tags, are analyzed with Spejd - a tool for simultaneous morphosyntactic disambiguation and shallow parsing [19], with a number of rules crafted to recognize multiword opinion patterns and apply sentiment modifying operations.

The Spejd formalism is a cascade of regular grammars. Unlike in the case of other shallow parsing formalisms, the rules of the grammar allow for explicit morphosyntactic disambiguation, independently or in connection with structure building statements, which facilitates the task of the shallow parsing of ambiguous and/or erroneous input.

For the purpose of sentiment analysis we extended the default Spejd's morphosyntactic tagset with a sentiment category expressing properties of positive or negative sentiment. This hybrid approach has been called Sentipejd [20].

Sentiment rules, discussed more extensively in [20], included (but were not limited to) the following:

- Affirmation - an expression of positive sentiment, usually an adverb confirming the sentiment of a positive word and should be treated as strong indications of sentiment (eg. 'I strongly recommend')
- Negation - as simple as the difference between "polecam" ('I recommend') and "nie polecam" ('I do not recommend'). The example generic rule captures also statements including the optional verb 'to be' ([base by]), like "nie jest dobry" ('isn't good').
- Nullification - expressing lack of a certain quality or property (usually of negative sentiment), for example "nie mam zastrzezen" ('I have no objections').
- Limitation - a limiting expression tells us that an expression of positive and negative sentiment has only a very limited extend, therefore hinting that the general sentiment of the review is the opposite of the expression. Example: "jedyny problem" ('the only problem').
- Negative modification - an adjective of negative sentiment preceeding a positive noun, for example "koszmarna jakosc" ('nightmarish quality').

Sentipejd returns either vectors of two integers (emoi=[pos, neg]) which express separately strengths of positive and negative emotions (it's not a simple sum of all emotional phrases) or the single, composite value – $\tau(c^{u_i}_{t_m})$. Every comment present in the collected dataset has been analyzed separately and the result has been stored as a vector in a database together with a category of the comment and the comment itself.

4.2 Reclassification Precision and the Emotional Distance

Although a similar natural language processing module has been already applied by authors to a broad variety of subjects (e.g. dynamic of public opinion[5]) the very first question which arises is: can a NLP system extract and evaluate emotions from usually very short and not always correctly (grammatical mistakes and typos) written comments? To answer this question, which is crucial for further deliberations, a standard data mining approach was used.

2 www.zetema.pl

Four separated, balanced subsets of comments were created:

- Set I (POS; NEG) - contains 2590 comments whereof 1295 are negative and 1295 positive,
- Set II (NEG; NEU) - contains 1454 comments whereof 727 are negative and 727 neutral,
- Set III (NEU; POS) - contains 1454 comments whereof 727 are neutral and 727 positive,
- Set IV (POS; NEU; NEG) - contains 2181 comments whereof 727 are positive, 727 negative and 727 neutral,

Every set of comments has been partitioned on testing and training set (30% and 70% cases respectively). For every set of comments three different classification approaches were used: neural network, support vector machine and decision trees (CHAID algorithm). As a target variable the label given by comment's author (negative, neutral or positive) was selected and the emo vector as input variables.

Table 4. Classification accuracy for different algorithms and testing subsets (average of four runs)

	Neural Network	Support Machine	Vector	Decision (CHAID)	Trees
POS and NEG (two classes; set I)	90,86%	90.36%		89,37%	
POS, NEU and NEG (three classes; set IV)	61,58%	61,66%		61,79%	
NEG and NEU (two classes; set II)	65,16%	65,80%		64,11%	
NEU and POS (two classes; set III)	71,15%	69,15%		70,24%	

The obtained results are presented in table 4. The first experiment was conducted to check if an evaluation based on the emotions expressed in comments and measured by the Sentipejd allows to predict the polarity of an label given by a human. At the beginning, the simplest subset was tested (only two classes - positive and negative - which should be relatively easier to separate). For the first set neural network approach was the most efficient. Over 90% classification accuracy indicates that the Sentipejd deals quite well with extracting emotions from texts (even not 100% correctly written) and that the significant difference in emotional content between positive and the negative labelled comments can be confirmed and measured.

Similar results for the neural network, support vector machines and decision trees (90.86%, 90.36% and 89.37% respectively) suggest that the reason for wrong classification goes beyond the classification algorithms. Only slightly better results for the same algorithms but validated on training sets instead of test sets seem to confirm that as well. A closer look at the misclassified cases shows that they belong into three (not always distinct) groups:

- written in a very specific slang, many misspellings, grammatical and orthographical errors, a lot of emoticons,
- well written but based on ironic, quizzical description of the past transaction,
- marked by user as positive but containing a negative evaluation,

The existence of the third group seems to confirm the results presented by Botsch and Luckner [7]. Some users, instead of leaving a negative mark, prefer to describe all the experienced problems in words. Because of their incoherency, those cases cannot be correctly classified using the adopted approach and they should be removed from the database. A detailed estimation of the scale of this effect requires manual processing of every comment which is not feasible because of the database size (1.7 million comments) and extends beyond the scope of this paper, although a rough estimation indicates that the effect of incoherent feedbacks is lower than 0.1% of all positive comments.

The biggest fraction of wrongly classified comments belongs to the first group. Many users, not only in e-Commerce systems but also on online forums, use a lot of abbreviations, emoticons, colloquial words and even intentionally misspelled words. Frequently, using intentionally transformed words is a sign of being a member of a specific social group. It helps users to identify the newcomers in an environment where cheap pseudonyms are present (a detailed study of the effects introduced by using cheap pseudonyms can be found here[16]). Some problems can be resolved (e.g. using a spell-checker to correct orthographical mistakes or creating a dedicated dictionary containing slang and colloquial words) but in principle intentional modifications of meaning or detecting irony will always be a challenge for computational linguistics.

The results for set IV are presented in the second row in table 4 Introduction of the third class made the task much more difficult. The results over 60% are still almost 30% better than in the baseline of random choice but significantly lower than for two classes. Thus, to check which comments cause problems for the classification algorithms, two more experiments have been conducted. Firstly, the separability for neutral and negative comments has been tested. The third row in table 4 contains the results for the set II which includes only negative and neutral comments. The classification precision slightly over 65% indicates that the emotional distance between neutrally and negatively tagged feedback is relatively small. Secondly, the same approach has been used to measure the emotional distance between neutral and positive comments. The results for all classification methods except support vector machines are at least 6% better and indicate that neutral comments are emotionally closer to negative.

To confirm the hypothesis stated in the previous paragraph a new testing set has been created. All the collected comments were split into two classes: one containing only positive labelled comments and one with negative and neutral feedback. Based on the emo vector (defined at the beginning of this section) and using the classification algorithms (support vector machines, neural network and decision trees) an attempt to rediscover the new classification has been done. The obtained results are slightly less precise than for the set I (positive and negative comments only; without neutrals) but the difference is about 3%. Thus, in most applications negative and neutral comments can be interpreted in the same way - as an expression of dissatisfaction. The label should not be treated as a scale of the experiences) because there is very little data to confirm the hypothesis that neutral feedback is less effective than negative.

4.3 Spiral of Hatred

The spiral of hatred is a well-known phenomenon present in a wide area of scientific fields (eg. Wydra identifies it as an core component of the war conflicts[22]) manifested as an endless action-reaction response, where successive iterations are subject to more negative emotion. Typically, in practical terms this effect can be observed on online forums where an initial misunderstanding causes a lasting exchange of messages containing many abusive words. Because reputation influences profitability of the seller[11] and every negative comment undermines this reputation, thus the reaction of a seller after receiving negative feedback can be more emotional. In fact, as a consequence of unbalanced levels of positive and negative comments, a very interesting heuristic has emerged. For experienced users, a single negative comment plays much more important role in the estimation of transaction risk that even many positive comments.

Because comments are visible after they are left, a natural place to express (and observe), the spiral-of-hatred effect is the reciprocal feedback given by the second party after transaction is completed. Thus, the database described in section 3.1. has been used to verify the spiral-of-hatred effect, which – referring to the formalism defined in section 3.2 - can be expressed as:

$$\forall \rho(c_{tm}^{u_i}), \rho(c_{tm}^{u_j}) \in \{NEG, NEU\} : \varphi(c_{tm}^{u_i}) < \varphi(c_{tm}^{u_j}) \rightarrow \tau(c_{tm}^{u_i}) > \tau(c_{tm}^{u_j}) \tag{1}$$

It is reasonable to assume (because of the sociological nature of the analyzed effect) that the above definition will not apply universally and to every single case. Therefore, in the first stage a weaker assumption was tested – the average of negative emotion for the second comment is higher than for the first:

$$\forall w, u \in U; s, t \in T : (\rho(c_s^u), \rho(c_t^w) \neq POS : \varphi(c_t^w) < \varphi(c_s^u)) \rightarrow \sum_{w,t} \tau(c_t^w) < \sum_{u,s} \tau(c_s^u) \tag{2}$$

The results are equivocal. First, the average value of τ for the comments given first is -0.63. The same value for the answers is higher and amounts -0.72. The difference is statistically significant at the level 0.07 which is a little bit above a typical 0.05 but it seems to make a spiral of hatred hypotheses at least very probable. Second, the standard deviations for both sets are almost equal – 2.01 – and it indicates that the distribution of emotion intensity between both the earlier and latter comment groups is similar but shifted. On the other hand, dividing the set analyzed in the previous paragraph into buyer and seller roles of the agent, makes results more complicated. More detailed results are presented in table 5.

Table 5. Average sentiment for buyers and sellers

	first	second
buyer	-0.55	-0.69
seller	-1.28	-0.96

In general, sellers are more emotional and more expressive than buyers ($\tau = -0.76$ as compared to $\tau = -0.60$ for buyers) and this pattern concerns both specific cases analyzed in table 3. There are at least three hypotheses which can explain this difference. Firstly, sellers write more correctly so the Sentipejd has an easier job extracting emotions from comments. Secondly, the cost of receiving negative evaluations for sellers is much higher (pseudonyms are more expensive and lower reputation affects profitability) and therefore boosts their reaction. Thirdly, sellers are simply more experienced and know how to make comments in a more negative way. As the standard deviation for sellers is only slightly higher than for buyers (should be significantly higher, if the source of difference in the emotional strength is related to misspellings and errors in comments), the second and the third hypotheses are the more probable ones.

The classical definition of the spiral-of-hatred effect formalized in eq. 1 and 2 is satisfied (and statistically significant) only for typical cases where buyers leave comments first. As expected, the average answer given by a seller is more negative. However, the same assumption is not true for the uncommon situation where sellers comment first. In that case the average negative sentiment in ordered pairs $\omega_{t_m} = (c_{t_m}^{u_i}, c_{t_m}^{u_j})$ such as $r(c_{t_m}^{u_i}) = seller$ is -1.12 and is much higher (-0.96) than for $r(c_{t_m}^{u_j}) = buyer$. The increase in negative emotions, compare to situation where buyers comment first, is observed symmetrically for both participants (buyers and sellers). Even though buyers answer very aggressively, at the end the emotional war is always won by sellers. They have stronger motivation because the reputation affects their profitability and are more experienced due to the extensive usage of the auction website.

More studies are needed to determine how the communication beyond auction platforms' cross-comments mechanism (e.g. via e-mail) influences emotional attitudes. However, on the very basic level the spiral-of-hatred effect can be identified in the collected data despite complex interactions of many social processes.

5 Discusion

Originally, the auction houses have been developed as goods exchange platforms where everyone could be either a seller or a buyer and where such roles are volatile and adopted only for one transaction. Nowadays, the auction platforms remind more of a shopping mall rather than a medieval bazaar and almost all members have clearly defined typical roles of either sellers or buyers. Therefore, it is necessary to revise the previous paradigm which used to determine the development of the reputation management systems. Instead of two more or less equal transaction parties, there is an explicit distinction: on one hand, sellers become more experienced due to the extensive usage of the auction system, on the other hand buyers' profitability is less sensitive to negative feedback.

The modification of the reputation system should take into consideration these facts. One of the possible ways to take them into account is for example to limit the possibility of leaving an evaluation by making it available only for buyers. One-sided comments make sellers defenceless, but elimination of negative reciprocal feedback will increase the likelihood that buyers comment more honestly. As an undesirable side effect of such a situation, blacklists of dishonest buyers can be created and maintained outside auction platforms, which can in turn be used as a tool for sometimes unjustified discrimination. More side effects should also be expected.

Another way to eradicate the spiral-of-hatred effect, which requires merely a minor modification in the existing reputation management systems, is to hold back the publication of an evaluation until an answer is sent. It should permit the elimination of the threat of revenge and thus make all comments more honest and less biased by the previous evaluation (more an answer based only on transaction experiences than an evaluation of the other participant performance). The problem that users will intentionally block publication of the negative comments can be solved by introducing a moderator who will be responsible for making an opinion visible (upon a request of one of the transaction parties) even if the answer does not appear. Even fewer changes are required to reduce the identified effect through establishing the minimum time-span that has to pass between comment and answer. Answers given right after a negative comment is received are more emotional and usually less informative.

Natural language processing tools are the best solution to investigate problems referred to in the descriptive part of submitted comments. Automatic sentiment extraction helps identify emotional wars immediately after they appear and either inform the administrators or even take appropriate steps automatically.. Analysis of every pair of comments can be complemented by the knowledge about typical behaviour of users taking part in transaction on the basis of their previous evaluations. Moreover, an efficient NLP algorithm can detect many discrimination strategies such as using a multitude of fake pseudonyms or atypical positive evaluations.

6 Conclusion

The broad variety of effects identified and described in this paper is only a fraction of all effects in auction websites. Jointly, with the stoning, slipping, self-selection[9], cheap pseudonyms[16], asymmetrical impact of positive and negative comments[11], price-reputation correlation[12], the importance of missing feedback[1], the presented results provide environment for invention, development and implementation of new techniques and tools with a goal to further increase satisfaction and usability of an auction website. Proposed changes can impact not only users' satisfaction but also profitability of the auction website.

The complex relationships between different users' behavioural patterns and hardly predictable side-effects discourage the managers responsible for maintaining and developing e-commerce systems from modifying the existing, proved solutions. They tend to use simple financial instruments like insurances or escrows to increase the level of security. Thus, the attempts to popularise the results collected by researchers over the last few years should be focused on the development of dedicated external tools to support users using those systems rather than on the modification of existing e-commerce systems.

Future research should be oriented toward sensitivity analysis of identified effects and influence of cultural circles and individual characteristics on the dynamics and existence of particular effects. Also, forecasting of social acceptance and social effects of the planned changes in an auction house is a challenging task[18]. Successful modeling and forecasting social responses (i.e. emergent attractors, stability points, non-linear dynamics) will be crucial to implement changes in Web 2.0 services.

References

1. Morzy, M., Wierzbicki, A.: The Sound of Silence: Mining Implicit Feedbacks to Compute Reputation. In: Spirakis, P.G., Mavronicolas, M., Kontogiannis, S.C. (eds.) WINE 2006. LNCS, vol. 4286, pp. 365–376. Springer, Heidelberg (2006)
2. Gregg, D.G., Scott, J.: A Typology of Complaints About Ebay Sellers. Communications of The ACM 51(4) (April 2008)
3. Lucking- Reiley, D., Bryan, D., Prasa, N., Reeves, D.: Pennies from eBay: The Determinants of Price in Online Auctio ns (1999), `http://eller.arizona.edu/~reiley/papers/PenniesFromEBay.pdf`
4. Bajari, P., Hortacsu, A.: Winner's Curse, Reserve Prices and Endogenous entry: Empirical Insights from eBay Auctions. Stanford Institute for Economic Policy Research. SIEPR. Policy paper No. 99-23 (1999)
5. Wawer, A., Nielek, R.: Application of Automated Sentiment Extraction from Text to Modeling of Public Opinion Dynamics. Journal of Environmental Studies 17(3B), 508–513 (2008) (in Polish)
6. Resnick, P., Zeckhauser, R.: Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. The Economics of the Internet and E-Commerce 11, S127–S157 (2002)
7. Botsch, J., Luckner, S.: Empirische Analyse von Bewertungskommentaren des Reputationssystems von eBay. Multikonferenz Wirtschaftsinformatik (2008) [DBLP:conf/mkwi/BotschL08]
8. Nam sued for libel over comments on eBay, `http://www.telegraph.co.uk/news/uknews/3247683/Man-sued-for-libel-over-comments-on-eBay.html`
9. Khopkar, T., Li, X., Resnick, P.: Self-selection, slipping, salvaging, slacking, and stoning: the impacts of negative feedback at eBay. In: ACM Conference on Electronic Commerce 2005, pp. 223–231 (2005) (DBLP:conf/sigecom/KhopkarLR05)
10. Lin, Z., Li, J.: The online auction market in China: a comparative study between Taobao and eBay. In: ICEC 2005, pp. 123–129 (2005)
11. Standifird, S.S.: Reputation and E-commerce: eBay Auctions and the Asymmetrical Impact of Positive and Negative Ratings. Journal of Management 27(3), 279–295 (2001)
12. Lee, Z., Im, I., Lee, S.J.: The Effect of Negative Buyer Feedback on Prices in Internet Auction Markets. In: Orlikowski, W., et al. (eds.) The Proceedings of the 21st International Conference on Information Systems, pp. 286–287 (2000)
13. Ozakca, M., Lim, Y.: A study of reviews and ratings on the internet. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, Montréal, Québec, Canada, April 22 - 27, 2006, pp. 1181–1186. ACM, New York (2006), `http://doi.acm.org/10.1145/1125451.1125673`
14. Scheff, T., Fearon Jr., D.: Social and Emotional Components in Self-Esteem. Journal of the Theory of Social Behavior 34, 73–90 (2004)
15. Ganesan, K.A., Sundaresan, N., Deo, H.: Mining tag clouds and emoticons behind community feedback. In: Proceeding of the 17th international Conference on World Wide Web, WWW 2008, Beijing, China, April 21 - 25, 2008. ACM, New York (2008)
16. Friedman, E.J., Resnick, P.: The Social Cost of Cheap Pseudonyms. Journal of Economics and Management Strategy 10, 173–199 (2001)

17. Talwar, A., Jurca, R., Faltings, B.: Understanding user behavior in online feedback reporting. In: Proceedings of the 8th ACM Conference on Electronic Commerce, EC 2007, pp. 134–142. ACM, New York (2007)
18. Kleinberg, J.: The convergence of social and technological networks. Commun. ACM 51(11), 66–72 (2008)
19. Buczynski, A., Przepiorkowski, A.: Demo: An Open Source Tool for Partial Parsing and Morphosyntactic Disambiguation. In: Proceedings of LREC 2008 (2008)
20. Buczynski, A., Wawer, A.: Shallow parsing in sentiment analysis of product reviews. In: Proceedings of the Partial Parsing workshop at LREC 2008, pp. 14–18 (2008)
21. Philip, J., et al.: The General Inquirer: A Computer Approach to Content Analysis. MIT Press, Cambridge (1966)
22. Wydra, H.: The Recurrence of Violence. Sociology Compass 2(1), 183–194 (2008)