Achieving Honest Ratings with Reputation-based Fines in Electronic Markets

Thanasis G. Papaioannou
Department of Computer Science,
Athens University of Economics and Business,
76 Patision Str., Athens, GR 10434, Greece
Email: pathan@aueb.gr

George D. Stamoulis
Department of Computer Science,
Athens University of Economics and Business,
76 Patision Str., Athens, GR 10434, Greece
Email: gstamoul@aueb.gr

***Abstract*—The effectiveness of online feedback mechanisms for rating the performance of providers in electronic markets is vulnerable to the submission of dishonest ratings. In this paper, we deal with how to elicit honest such ratings in a competitive electronic market where each participant can occasionally act both as provider and as client. We assume that each service provision is rated by both parties involved; only upon agreement, this rating is included in the calculation of reputation for the provider's performance. We first study as a single-shot game the effectiveness of inducing, upon evidence of lying (i.e. disagreement of the submitted feedback), fixed fines to both transacted parties, yet different ones for the provider and the client. We prove that the submission of honest feedback can be a stable equilibrium for the whole market under certain initial system conditions. Then, we refine our game-model for repeated transactions and calculate proper different reputation-based fines for lying. These fines enable the submission of honest feedback as a stable Nash equilibrium of the repeated game and reduce the social losses due to unfair punishments. Finally, we argue that our model is appropriate for analyzing actual electronic markets, and we investigate the impact of employing our approach to the feedback schemes of such markets.**

I. Introduction

Electronic markets, such as eBay (http://www.ebay.com), have already become very popular for trading nearly all kinds of products and services. Such markets involve hidden information on the quality of services and the honesty of the provider. Due to this information asymmetry, there is a possibly high risk for a client to get from a transaction lower value than expected. It has been established by Dellarocas in [1] that reputation can be a proper means of revealing the hidden information on low-performing participants in online markets and in peer-to-peer systems that do not involve payments. For the latter systems, we have shown in [2] that the calculation of reputation values has to be complemented by proper reputation-based policies that determine the pairs of peers eligible to interact with each other. However, the accuracy of reputation values is based on the honesty of the feedback on the quality of provided services. False or strategic feedback may arise, especially if such rating behavior provides some value to the rater. We have introduced and analyzed in [3] a very effective mechanism that provides the incentives for truthful rating in a peer-to-peer system with a dynamically evolving population, isolating large fractions of liars from the system. According to this mechanism, both transacting peers (rather than just the client) submit ratings on the performance/quality of their mutual transaction. If ratings disagree, then both transacting peers are punished, i.e. isolated from the system for a time period. The motivation for this is that such a disagreement is a sign that one of the transacting peers is lying, but the system cannot tell *which one*, due to hidden information. On the other hand, if two transacting peers agree in their ratings, then this feedback is taken into account in the calculation of the provider's reputation.

In this paper, we deal with the problem of promoting the submission of truthful ratings in competitive online markets with payments, where each participant can act both as provider and as client. We made a preliminary attempt to address this problem in [4]. There are many interesting service examples that match this context, e.g. exchange (in the sense of selling and buying) of software modules among programmers, of news among agencies, or even of collectible goods such as vinyl records of classic music, stamps etc. The present context is different from that of peer-to-peer systems considered in [3] and the problem of promoting honest feedback is now more complicated due to the payments. That is, it is plausible that either the client or the provider has a direct monetary benefit from submitting a strategic rating and/or a future competitive advantage in the market for providing services. Thus, we adapt properly the incentive mechanism of [3], so as to be applicable in the new context. In particular, we propose that disagreement induces direct monetary penalties (i.e., fines) to both transacted parties; this gives rise to interesting trade-offs that also relate to the future positioning of the two parties in the market. Moreover, in order to balance the risks due to hidden information of both parties of a transaction (i.e. a service- or a product- provision), the client pays ahead a certain portion of the total payment, while the remaining amount is *only* paid if both parties agree that this transaction was successful.

The contribution of this paper is as follows: We develop and analyze a game-theoretic model capturing the effects of the modified incentive mechanism for honest feedback in the context of interest. We establish that employing proper fixed fines (yet different ones for the provider and for the client of each transaction) for disagreement, we can enforce a stable equilibrium with honest feedback in the market under certain

conditions, which we thoroughly investigate. Moreover, we calculate proper non-fixed reputation-based fines that render honest feedback a Nash equilibrium, which is experimentally proved to be stable. Reputation-based fines reduce the social welfare losses due to unfairly induced fines to providers and clients. Our results apply even if a participant employs a different account for each role. Finally, we investigate the impact of employing our approach to eBay.

II. Related Work

Apart from [3] that has already been discussed, there is significant work that is related to incentives for honest rating. In [5], Dellarocas deals with truthful reporting in online trading communities where collaborated liars constitute at most 10% of the entire population of buyers; unfair ratings and discriminatory behavior are dealt with by clustering the ratings of buyers with similar tastes based the rating they submit on the transactions with sellers they have in case. A different mechanism for truthful reporting of the provided quality is proposed by Dellarocas in [6], where providers publish their services and their promised quality in a market list. The mechanism of [6] publishes a modified estimate of the expected quality for the next service instance of a provider, so as to compensate for the payoff gains resulted by deviations between the promised and the offered quality levels by the provider in the previous service instance. In case of sincere clients, honest rating by providers is a subgame perfect equilibrium.

A side payment approach for eliciting honest feedback in electronic markets is proposed in [7] by Miller *et al.* In particular, a payment charged to the current buyer is paid to the subsequent buyer transacting with the same seller, according to a rule for scoring her prediction of the rating of a later buyer for their common seller. This approach enables truthful reporting of clients to be a Nash equilibrium, yet not a unique one. Another side payment mechanism is proposed by Jurca and Faltings in [8] for truthful reporting on the performance of a hotel in the online hotel booking industry. The occupancy of the hotel is a linear function of its reputation. The hotel decides on the quality of service offered, which depends on investment. After a service provision, the hotel reports whether having offered high quality or not. Then, the client decides on agreeing or disagreeing with the report of the hotel, or on leaving. This situation corresponds to a Stackelberg game. In case of disagreement, different fixed punishment fees ϵ_H and ϵ_C, with $\epsilon_H > \epsilon_C$, are charged to the hotel and the client respectively, and a negative vote for the hotel is counted. Otherwise, a positive vote is counted for the hotel. It is proved that when the hotel and the clients are rational, the percentage of false reports does not exceed a certain threshold. Also, it is proved that when there is employed reputation for truthful reporting and clients are long-lived, then the number of cheats against the same client are bounded. Strategic reporting by clients was not considered in [8], which is a major consideration in our work both in [3] and in this paper.

In this paper, we study several issues on truthful reporting of ratings' feedback that are either innovative or not covered adequately by the aforementioned works: a) The stability properties of the honest-rating Nash equilibrium that is enforced by the proposed mechanism, b) the dual role of participants in the market, and c) the reputation-based fines for untruthful reporting.

III. The Basic Model

We consider an e-marketplace with many participants where each of them can act both as provider and as client of a certain service (or product). Next, we present our assumptions to model such a market. Time is discretized in rounds. At the beginning of each round, each of the N participants decides whether she will act as a provider (with probability q) or as a client (with probability 1-q), where q is bounded away both from 0 and 1. Then, a random client selects a provider of the requested service to transact with. The size N of the population is taken to be large and fixed, although the population may be dynamically renewed. An instance of the service (or the product) exchanged generates a fixed utility u to the client regardless of who was the provider, yet under the condition that the service was provided successfully, i.e. with quality above a pre-specified level. Each successful service provision requires that the provider exerts effort with cost v. The participants of the market may belong to different performance types. In particular, associated with each participant i is a success probability a_i, which is private information known only to himself.

A reputation metric for performance is employed in order to reveal performance types of the participants. After each service provision, *both* participants are expected to submit feedback to the reputation system on the outcome of the service. This feedback includes a binary characterization of the service (i.e. "successful" or "unsuccessful"). Whenever meaningful, this is combined with a measurable parameter, in order to justify the binary rating; e.g. in the case of trading storage capacity the participants that transacted may also report the quantity of GBytes provided. If the two feedback reports are indeed in agreement, then the corresponding binary rating is aggregated into the reputation value of the provider. This is attained by employing an appropriate aggregation function introduced in [4]; see Section V. If either the feedback reports are not consistent or exactly one of the parties does not report feedback, then monetary penalties are imposed to both parties. Recall that disagreement in feedback reports is an evidence of lying, and thus it invokes a punishment. If only one transacted party sends feedback, then the reputation system cannot decipher whether this transaction actually happened. Thus, if exactly one feedback report is received for a hypothetical transaction, then both parties are charged disagreement fines. This feature provides the right incentives against abnormal promotion or demotion of others. Note that due to the inclusion of the measurable parameter in the feedback report, it is hard for the participants that transacted to have agreed either by chance or strategically due to *temporary* collusion. The implementation

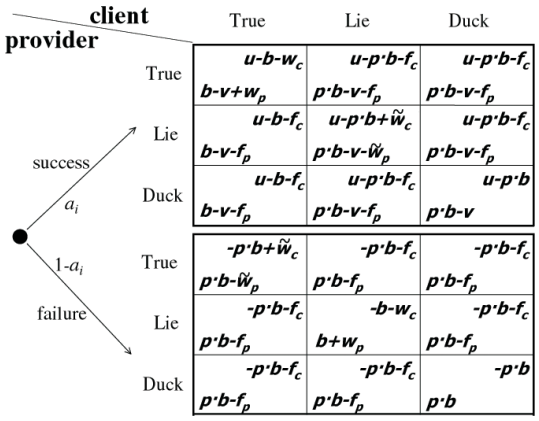


Fig. 1. The single-shot game.

of this mechanism requires certain observing authorities and can be either centralized or distributed; see [3] for a detailed presentation of the related implementation issues. The fines can be enforced if participants have to pay them in order to stay or re-enter the market after a disagreement.

Furthermore, each service instance is charged at a fixed price b in all rounds. This price is set by the market and does not depend on the specific participants that transact. Thus, due to hidden information, the following risks arise in a service provision: a) the client may receive a service of unsatisfactory quality, while b) the provider may provide a service successfully and not be paid by the client. In order to balance these risks the client *prepays* a certain portion $p \cdot b$ of the total price b, while the remaining amount $(1-p) \cdot b$ is paid only if both parties agree that this transaction was successful. Refusal of post-payment by the client is necessarily accompanied by a negative rating.

In order for the right incentives for performance to be provided to participants we assume that a reputation-based *policy* for provider selection is employed [2]. In particular, in each round, clients are assumed to associate to providers in a probabilistically fair manner according to the reputation values of the latter [2]. That is, if participant i serves at a certain round as a provider, then the probability that she will be selected by a participant j who acts now as a client is proportional to r_i. Another provider selection policy analyzed in [2] is Max-Reputation, which prescribes that each client simply selects the provider with the highest reputation. If, however, we assume that only a uniformly picked random subset of the providers can offer the service requested by the client of each particular round, then Max-Reputation should be restricted to the appropriate subset of providers. Under this assumption, the demand attracted by a provider tends to approximate that of the assumed probabilistically fair reputation-based selection policy. Furthermore, say that in a certain round, the set of participants to act as clients (resp. providers) is C (resp. P). Note that the expected size of the set C of clients is $E[|C|] = N(1-q)$, while that of P is $E[|P|] = Nq$. Since N is large and q is bounded away from 0 and 1, both sets of providers and clients are expected to have a large number of members selected randomly. Therefore, we have $\frac{1}{|P|}\sum_{k\in P} r_k \approx \frac{1}{N}\sum_{k=1}^{N} r_k = \bar{r}$, which implies that $\sum_{k\in P} r_k \approx \bar{r}|P| \approx \bar{r}Nq$, where $\bar{r}$ is the mean reputation value over all participants in the market. Furthermore, notice that the probability for a certain provider i to be selected by a certain client j is $\frac{r_i}{\sum_{k\in P} r_k}$. Using the previous approximation, it follows that this probability equals approximately to $R_i/(Nq)$, where $R_i = \frac{r_i}{\bar{r}}$ is referred to as *rank* of participant i. Hence, the demand attracted by provider i depends on her rank R_i, i.e. on her reputation r_i normalized by the mean reputation $\bar{r}$, rather than solely on r_i.

In Figure 1, we present the extensive form of the single-shot game that is played in each round by the participant transacting as a client with the provider she has selected. This game consists of two sub-games, namely that for the case of success of the service provided and that for the case of failure. Clearly, the parties involved have three alternative pure reporting strategies, namely those in S= {*True*, *Lie*, *Duck*}. Thus, they can choose either to submit feedback truthfully or to submit false feedback or not to send feedback at all. If the players' feedback ratings are in disagreement, or if exactly one of them does not submit feedback, then no rating is taken into account, but the provider and the client incur monetary penalties f_p and f_c respectively. Initially, we have take that these fines are fixed. This restriction is relaxed in Section V.

On the other hand, in case of agreement, the common rating is taken into account and the provider's reputation is updated. Thus, due to the reputation-based policy described above, each agreed rating affects the expected future payoffs of the players. In particular, the impact of a positive rating in the expected future payoff of a provider is denoted as w_p, which is positive due to the fact that the provider's rank increases, thus enabling her to attract more clients in the future. Also, the impact of a negative rating is denoted as $-\tilde{w}_p$, which is negative becuase the provider's rank decreases. The payoff of a client is affected too by an agreed rating on the provider's performance, yet in a reverse way. The rating influences the mean reputation $\bar{r}$, and in turn the client's rank, which determines the demand that she will attract whenever she acts as a provider in future rounds. The payoff impact for a client of an agreed positive rating is denoted as $-w_c$, which is negative, because the mean reputation $\bar{r}$ increases and thus the client's rank decreases; also the impact of a negative rating is denoted as $\tilde{w}_c$, which is positive as the client's rank increases. The payoff impacts $w_p, -\tilde{w}_p, -w_c, \tilde{w}_c$ are initially taken as fixed, i.e. independent of the provider's and the client's rank values, which as already explained influence the respective demand that they will attract in the future. This assumption is relaxed in Section V. It can be seen that a system where the above payoff impacts are indeed almost fixed is a large market employing a reputation measure that equals the difference between successful and unsuccessful service provisions for each participant. Indeed, for this system, the increase (resp. decrease) of the rank of a provider following a successful (resp. an unsuccessful)

service provision is approximately $1/\bar{r}$ (resp. $-1/\bar{r}$), which is independent of her previous rank; also, the payoff impact for the client is approximately equal to 0.

As already explained, all terms of the payoff matrix of the game presented in Figure 1 are *independent* of the players' reputation values. Moreover, each of the players decides on her strategy after observing the outcome of the service provision, which is much more revealing information than the reputation of the provider. Therefore, the choice of strategy of each of the players in the game is influenced neither by her own reputation nor by that of her opponent.

IV. Stability Conditions for Truthful Equilibrium

In this section, our goal is to derive necessary conditions under which truthful reporting by participants in both roles is a stable Nash equilibrium in both the success and the failure subgames of Figure 1. Note that we are interested in finding the lowest values of f_p and f_c that satisfy our objectives, in order to limit the associated social loss. Indeed, an unnecessarily high punishment could serve as credible threat against lying but it would also lead participants to leave from the e-marketplace. Therefore, we first employ equilibrium analysis for each of the two subgames of Figure 1, and derive conditions for truthful reporting to be best response to itself both for providers and for clients regardless of the outcome of the transaction. Thus, under these conditions, truthful reporting is a Nash equilibrium in both subgames of Figure 1. Then, we employ evolutionary game theory [9], in order to study the stability of the desired Nash equilibrium.

Paralleling part of the analysis of [8], we first note that if the participants could be assumed to be truthful when in the client role , then it would be a *dominant* strategy for them to be truthful as providers when succeeding in providing their services. Moreover, in when failing in their service provision, truthful reporting can again be the dominant strategy for providers, if the monetary punishment (i.e., fine) that is charged to providers in case of disagreement is $f_p > \tilde{w_p}$. Thus, when clients are truthful, providers can be made truthful too. The condition $f_p > \tilde{w_p}$ represents the lowest bound for any fine that has to be induced to providers in order for them to always submit truthful feedback. On the other hand, if providers could be assumed to be truthful, then it would be a dominant strategy for clients to be truthful as well when providers fail in service provision. When providers succeed, then the disagreement fine should satisfy the condition $f_c > b(1-p) + w_c$, in order for truthful reporting to be dominant strategy for clients too. Summarizing the above analysis, $f_c > b(1-p) + w_c$ and $f_p > \tilde{w_p}$ are sufficient conditions for *True* to be best response to itself for both providers and clients, which implies that [*True*, *True*] is a Nash equilibium in both subgames. These conditions are henceforth assumed. Note that there is no dependence of f_p, f_c on w_p and $\tilde{w_c}$. Indeed, for the success sub-game, w_p constitutes the provider's payoff impact of a positive vote; if the the client is truthful and thus submits such a vote, then the provider has no incentive to disagree since w_p has been naturally assumed to be positive. Thus, no condition on disagreement punishments can involve w_p. Similar reasoning applies for w_c.

The only reasonable assumption regarding reporting behavior is that both providers and clients act rationally so as to maximize their payoffs. Henceforth, we analyze our game-theoretic model under this assumption. First, note that there are no disagreement punishments f_p, f_c (and payoff impacts of votes w_p, w_c, $\tilde{w_p}$, $\tilde{w_c}$) that render *True* dominant strategy in either subgame of Figure 1. When $f_c > b(1-p) + w_c$ and $f_p > \tilde{w_p}$, there are three pure Nash equilibria and one in mixed strategies in each subgame: i) Both provider and client report truthfully. ii) Both provider and client lie. iii) Both provider and client do not submit feedback. iv) Mixed strategies for transacting parties in either subgame, which are denoted as $\vec{x_M}$, $\vec{y_M}$ for providers and clients respectively in the success subgame. The mixed strategy vectors in equilibrium define the respective probabilities that the pure strategies are employed by the players, so that if a player employs a mixed strategy then its opponent is indifferent among her pure strategies, i.e. the opponent's expected payoffs of all three pure strategies are equal to each other. Two important questions are in order: Which Nash equilibrium is ultimately selected by participants? Is this equilibrium stable? In order to answer these, we employ the theory for evolutionary stability in games [9]. According to this theory, an *Evolutionary Stable Strategy* (ESS) is a Nash equilibrium strategy that is also a better response to any mutant strategy played by a small fraction of invaders than the mutant strategy is to itself. A totally or partly dominated strategy cannot be ESS. However, since $f_c > b(1-p) + w_c$ and $f_p > \tilde{w_p}$, no dominated strategies exist. According to Proposition 2.4 of Samuelson [9], in either asymmetric subgame between providers and clients, only the three equilibria in pure strategy pairs (namely [*True*, *True*], [*Lie*, *Lie*], [*Duck*, *Duck*]) can be ESSs of the symmetric version of either subgame, because Nash equilibria in mixed strategies are not strict.

Next, we investigate the necessary conditions in order for the [*True*, *True*] strategy-pair to be selected by participants in either subgame, i.e. to be the ESS that the e-marketplace will evolve to. For this purpose, we study the relative fitnesses (i.e. expected payoffs) of strategies and their evolutionary dynamics [9]. That is, the per capita increase rate along time of the fraction of the participants that play according to each of the three strategies in either subgame. Thus, the evolutionary dynamics in the success subgame of Figure 1 are given by:

$$\begin{aligned} &\dot{x_1} = x_1(\pi_1 - \bar{\pi}),\ \dot{x_2} = x_2(\pi_2 - \bar{\pi}),\ \dot{x_3} = x_3(\pi_3 - \bar{\pi}), \\ &\dot{y_1} = y_1(\hat{\pi}_1 - \bar{\hat{\pi}}),\ \dot{y_2} = y_2(\hat{\pi}_2 - \bar{\hat{\pi}}),\ \dot{y_3} = y_3(\hat{\pi}_3 - \bar{\hat{\pi}}). \end{aligned} \quad (1)$$

1, 2, 3 stand for *True*, *Lie*, and *Duck* strategies respectively, and x_j (resp. y_j) for the fraction of the population playing according to strategy j when acting as provider (resp. client); also, π_j, $\hat{\pi}_j$ denote the expected payoffs for providers and clients respectively that play according to strategy j. For example, the expected payoff for a provider that plays *True* in the success subgame is given by $\pi_1 = y_1(b-v+w_p)+(y_2+y_3)(pb-v-c)$.

Also, $\bar{\pi}$, $\bar{\tilde{\pi}}$ are the average expected payoffs over the various pure strategies of providers and clients respectively in the success subgame. The equations (1) also apply to the failure subgames. We denote the corresponding variables with the same notation as in the success subgame followed by a "′".

Finally, the conditions on the fractions of the population that play according to the various strategies under which a strategy is ESS amounts to deriving this strategy's *basin of attraction*. In particular, the basin of attraction for a strategy-pair in the success subgame corresponds to a region for $(x_1, x_2, x_3, y_1, y_2, y_3)$ with the following property: If the system's population mix lies in this region at some point in time, then all providers and clients will asymptotically (in time) report according to the strategies of this pair. Of course, a different basin of attraction is associated with each of the three Nash equilibria in pure strategies. Since $x_1 = 1-x_2-x_3$ and $y_1 = 1-y_2-y_3$, each basin of attraction can be expressed by the Cartesian product of a region for (x_2, x_3) and one for (y_2, y_3). A similar property applies to the basins of attraction for the strategy-pairs in the failure subgame.

Theorem 1: Truthful reporting is the only ESS for all participants of the e-marketplace, iff the population fractions of providers and clients that play according to the various strategies are in $X^* \times Y^*$ in the success subgame and are in $X^{*\prime} \times Y^{*\prime}$ in the failure subgame where:

i) $(x_2,\, x_3) \in X^*$ iff $x_3 \leq \frac{f_c - b(1-p) - w_c}{2f_c - w_c}$ and $x_2 < \min\{\frac{f_c - b(1-p) - w_c - (f_c - w_c)x_3}{2f_c + \tilde{w_c} - w_c}, \frac{f_c - w_c - b(1-p) - (2f_c - w_c)x_3}{f_c - w_c}\}$,

ii) $(y_2,\, y_3) \in Y^*$ iff $y_3 \leq \frac{f_p + w_p}{2f_p + w_p}$ and $y_2 < \min\{\frac{f_p + w_p}{2f_p - \tilde{w_p} + w_p}(1 - y_3),\, 1 - y_3 \frac{2f_p + w_p}{f_p + w_p}\}$,

iii) $(x'_2,\, x'_3) \in X^{*\prime}$ iff $x'_3 \leq \frac{f_c + \tilde{w_c}}{2f_c + \tilde{w_c}}$ and $x'_2 < \min\{\frac{f_c + \tilde{w_c}}{2f_c - b(1-p) + \tilde{w_c} - w_c}(1 - x'_3),\, 1 - x'_3 \frac{2f_c + \tilde{w_c}}{f_c + \tilde{w_c}}\}$,

iv) $(y'_2,\, y'_3) \in Y^{*\prime}$ iff $y'_3 \leq \frac{f_p - \tilde{w_p}}{2f_p - \tilde{w_p}}$ and $y'_2 < \min\{\frac{f_p - \tilde{w_p}}{2f_p + b(1-p) - \tilde{w_p} + w_p}(1 - y'_3),\, 1 - y'_3 \frac{2f_p - \tilde{w_p}}{f_p - \tilde{w_p}}\}$.

Outline of Proof: Being a strict Nash equilibrium of this asymmetric game, the strategy-pair [*True*, *True*] is also an ESS, according to [9]. Within its basin of attraction, truthful reporting strategy should be more fit than the other two pure reporting strategies of providers and clients. Thus, in its basin of attraction, the per capita rate of growth of truthful reporting strategy should be greater than those of the other two strategies in both subgames and for both the providers and the clients. Therefore, the following inequalities should apply to $j = 2, 3$:

$$\frac{\dot{x_1}}{x_1} > \frac{\dot{x_j}}{x_j},\ \frac{\dot{y_1}}{y_1} > \frac{\dot{y_j}}{y_j},\ \frac{\dot{x'_1}}{x'_1} > \frac{\dot{x'_j}}{x'_j},\ \frac{\dot{y'_1}}{y'_1} > \frac{\dot{y'_j}}{y'_j}\,. \qquad (2)$$

Combining these with the expressions of evolutionary dynamics (1) and the fact that $x_2,\, y_2,\, x'_2,\, y'_2 \in [0, 1]$, we derive, after some algebra, the conditions on $(x_2,\, x_3)$ and $(y_2,\, y_3)$ that define $X^* \times Y^*$ and those on $(x'_2,\, x'_3)$ and $(y'_2,\, y'_3)$ that define $X^{*\prime} \times Y^{*\prime}$. ■

Note that the above theorem is very important as [*Lie*, *Lie*] and [*Duck*, *Duck*] strategy pairs are also strict Nash equilibria of the asymmetric game and thus ESSs. Their basins of attraction in either subgame are also Cartesian products of two quadrilaterals per ESS one for providers and one for clients. These are depicted in Figures 2(a) and 2(b) for the success subgame. In each figure, the corresponding triangle of feasible solutions (e.g, the triangle $x_2 \in [0, 1]$, $x_3 \in [0, 1]$ and $x_2 + x_3 \leq 1$ in Figure 2(a)) is partitioned into three basins of attraction, each corresponding to a pure strategy ESS, which are separated by bold boundary lines. For the success subgame, the basins of attraction of each of the three ESSs are the Cartesian product of their respective quadrilateral for provider in Figure 2(a) with its counterpart quadrilateral for clients in Figure 2(b). Similar properties apply to the failure subgame and the corresponding basins of attraction. Note also that the trajectories in the figures show possible evolution paths of the fractions of the population that play each of the three strategies with various starting points (i.e. initial conditions on these fractions). The point of intersection of the bold lines in each of the Figures 2(a) and 2(b) comprises the mixed strategy equilibrium in the success subgame, denoted as $\vec{x_M}$ and $\vec{y_M}$ for providers and clients respectively. Observe that although mixed strategy equilibria in the success and the failure subgame are stationary points, they are not ESSs. Indeed, any mutation from there leads participants to one of the Nash equilibria in pure strategies, in both subgames. Finally, note that it is possible that the fractions of providers belong to the basin of attraction of one ESS while the fractions of clients belong to the basin of attraction of another ESS. Then, according to evolutionary dynamics, these fractions will keep changing; thus, the population mix will move towards the boundaries between the areas of attraction of the two ESSs, until only one of them prevails for both player roles.

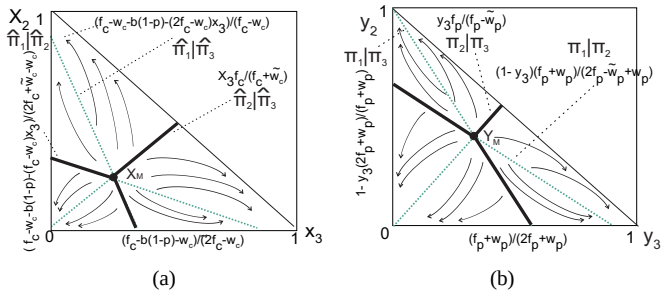


Fig. 2. The basins of attraction of all three ESSs (a) for providers and (b) for clients in the success subgame.

V. Extended Model: Reputation-based Payoff Impact and Fines

In this section, we extend the model of Section 3 by calculating the payoff impact resulted by a positive or negative vote for participants on the basis of reputation (essentially of the rank), and by analysing the resulting repeated game. To this end, ratings' feedback should be aggregated properly in order the history of a provider's performance to be summarized in a meaningful reputation metric. Moreover, if proper reputation-based policies [2] are employed, then reputation

provides incentives to providers for improved performance. We introduced in [4] an innovative reputation metric that is updated according to the formula $r' = \beta r + (1-\beta)\mathbf{1}(\text{success})$, where $\beta < 1$ is the discount factor for the past transactions. Besides being tractable, its main merit is that the reputation value does not depend explicitly on the number of service provisions already performed by the participant, contrary to the Beta metric [10]. In fact, this above aggregation formula provides reputation values equal to the numerator of Beta. Thus, in order to maintain a high reputation value (i.e. close to 1), a provider should keep on offering services successfully. When this new reputation metric is employed, the expected reputation value $E[r_i]$ after n service provisions for a participant i that provides services successfully with probability a_i is $E[r_i] = a_i(1-\beta^n)$. Observe that, as $n \to +\infty$, $\beta^n \to 0$, and the expected reputation value given by the new reputation metric equals that under Beta, namely a_i. The difference in the ranks of the transacted participants after a positive vote are denoted as $\Delta R_p^+(R_p)$, $\Delta R_c^+(R_c)$ for the provider with rank R_p and the client with rank R_c respectively. The respective differences after a negative vote are $\Delta R_p^-(R_p)$, $\Delta R_c^-(R_c)$. It appears that all these differences depend on the participant's rank, but yet they are independent of the number of services already offered by the provider. In particular, in the proof of Theorem 2 to follow, we employ the formula $\Delta R_p^-(R_p) = \frac{\beta R_p}{1-R_p\frac{1-\beta}{N}} - R_p$. Henceforth, we consider the single-shot game of Figure 1 played *repeatedly* by a fixed participant, yet with two important differences: a) In case of disagreement, the monetary punishment depends on the rank of the participant as well as on whether she acted as provider or client. b) In case of agreement, the future payoff-impacts $w_p, -\tilde{w_p}, w_c, -\tilde{w_c}$ are not included in the total payment as fixed terms. Instead, we quantify them precisely as follows: We take explicitly into account the impact of the agreed vote on the ranks of each of the two participants in their expected payoffs in the future rounds. We assume that the payoff of each extra future round is discounted by an extra factor $\delta < 1$. In fact, in the case of dynamically renewed population, $\delta = d(1-\frac{1}{T})$ where d is the discount factor and T is the mean of the geometrically distributed lifetime of each participant. Next, we derive reputation-based fines that render truthful reporting a Nash equilibrium of the repeated game. First, we compute the per round expected payoff of a participant assuming that all participants report truthfully. In particular, the expected per round payoff $V(R_i^{(t)}, a_i)$ of a participant i with rank $R_i^{(t)}$ and success probability a_i at round t is:

$$
\begin{aligned}
V(R_i^{(t)}, a_i) &= q\frac{1}{qN}R_i^{(t)}[a_i(b-v) + (1-a_i)pb] \\
&+ (1-q)\frac{1}{(1-q)N}[\bar{a}(u-b) + (1-\bar{a})(-pb)] \qquad (3) \\
&= \frac{1}{N}\{R_i^{(t)}[a_i(b-v) + (1-a_i)pb] + \bar{a}(u-b) - (1-\bar{a})pb\}\,.
\end{aligned}
$$

This expectation applies prior to the randomized decision of the participant regarding whether she will act as a provider or a client and also includes the cases where a) i acts as a client but she is not the one to transact at round t and b) i acts as a provider but she is not the one to be selected by the client to be served at round t. In equation (3), $\bar{a}$ is an estimate of the probability that the participant will be provided a successful service by the "average" provider (with respect to the reputation-based selection policy) if she acts as the client to be served in the round considered. This probability $\bar{a}$ can be taken as independent on the rank $R_i^{(t)}$ of participant i. Indeed, the provider selection policy is common for all clients while all clients are treated evenly by providers regardless of their ranks. $\bar{a}$ can be approximated by employing the distribution of the reputations of the various participants. (Its only dependence on the specific client lies in the fact that $\bar{a}$ is actually an average value over all the remaining potential providers, which is a negligible dependence in a large population.) However, such an estimate is actually unnecessary for the analysis to follow, as it is seen below. Note also that $R_i^{(t)}$ denotes the participant's rank prior to any action taken in round t. The impact of the transaction at t on the rank of participant i appears in $R_i^{(t+1)}$. This also applies to all other participants. Next, we derive the reputation-based punishments in the case of disagreement.

Theorem 2: Truthful reporting is a Nash equilibrium of the repeated game for all players in both roles of the game, if monetary punishments (i.e. fines) $f_p(R_i) > g(R_i)$ and $f_c(R_j) > \hat{g}(R_j)$ are charged in case of disagreement in feedback reports to transacted parties i (provider) and j (client) with ranks R_i and R_j respectively; $g(\cdot)$, $\hat{g}(\cdot)$ are functions of rank.

Outline of Proof: According to [9], truthful reporting is a Nash equilibrium of the repeated game if and only if deviating from this strategy *for a single round* results in a strictly smaller payoff for the deviant participant than continuous truthful reporting. Clearly, such a deviation is meaningful only for a provider that failed in service provision and only for a client that was offered a service successfully.

First, we consider a particular participant i that acted unsuccessfully as a provider at round t. Since the corresponding client is by assumption truthful, if i would either *Lie* or *Duck* at the end of round t, then she would incur a fine $f_p(R_i)$ while her rank $\tilde{R}_i^{(t+1)}$ at the next round would equal $R_i^{(t)}$, because due to the disagreement arising, there is no new vote to be aggregated with the previous reputation of participant i. On the other hand, if i reports ratings' feedback truthfully, then her future rank $R_i^{(t+1)}$ will be $R_i^{(t+1)} = R_i^{(t)} + \Delta R_p^- < R_i^{(t)}$. Thus, by being truthful, participant i avoids the immediate loss due to the monetary punishment but experiences a loss in her future payoff, due to the reduction of her rank. This loss equals $\sum_{\tau=t+1}^{\infty}\delta^{\tau-t}[V_i(\tilde{R}_i^{(\tau)}, a_i) - V(R_i^{(\tau)}, a_i)]$, where: $\{R_i^{(\tau)}\}_{\tau=t+1}^{\infty}$ (resp. $\{r_i^{(\tau)}\}_{\tau=t+1}^{\infty}$) is the sequence of future ranks (resp. reputations) of the participant i if she is truthful at round t and at all other rounds to follow, while $\{\tilde{R}_i^{(\tau)}\}_{\tau=t+1,\ldots}$ (resp. $\{\tilde{r}_i^{(\tau)}\}_{\tau=t+1,\ldots}$) is the sequence of future ranks (resp. reputations) of participant i if she is not truthful at round t (thus avoiding the reduction of her future rank) and she is

truthful at all other rounds to follow. Therefore, the deviation under discussion will not occur, as long as the punishment $f_p(R_i^{(t)})$ for the provider is large enough to exceed the loss in future payoffs, i.e. if and only if

$$f_p(R_i^{(t)}) > \sum_{\tau=t+1}^{\infty} \delta^{\tau-t}[V(\tilde{R}_i^{(\tau)}, a_i) - V(R_i^{(\tau)}, a_i)] \quad (4)$$

$$= \frac{1}{N} \cdot [a_i(b-v) + (1-a_i)pb] \sum_{\tau=t+1}^{\infty} \delta^{\tau-t} E[\tilde{R}_i^{(\tau)} - R_i^{(\tau)}] \,,$$

where we have also used equation (3). The expected difference $E[\tilde{R}_i^{(\tau)} - R_i^{(\tau)}]$ that arises in the respective ranks of the specific participant i at round τ due to her deviation at round t can be bounded approximately by *coupling* the two trajectories of the system for the rounds after t; namely, the trajectory arising in the case where participant i deviates from truthful reporting at round t and the one corresponding to the case where she is truthful in this round too. Thus, it follows after some algebra, that, for large N, the expected difference at some subsequent round τ in the reputation values of the participant i in the two trajectories can be bounded by $-\Delta R_p^-$, which is the corresponding difference at $\tau = t+1$. Although somewhat complicated to prove, this bound is intuitively clear: Since the only deviation of participant i from truthful reporting occurs at round t, the maximum expected deviation of future rankings of the two trajectories can only arise immediately, i.e. at round $t+1$. That is, $E[\tilde{R}_i^{\tau} - R_i^{(\tau)}] \leq -\Delta R_p^-$. Furthermore, notice that, since N is large, we can approximate the expression of $\Delta R_p^-(R_p) = \frac{\beta R_p}{1 - R_p \frac{1-\beta}{N}} - R_p$ by omitting the terms $\Theta(\frac{1}{N})$. However, the multiplicative factor $\frac{1}{N}$ is a dominant one in $V(\tilde{R}_i^{(\tau)}, a_i) - V(R_i^{(\tau)}, a_i)$ and therefore it cannot be omitted from there as well. Thus, we obtain $\Delta R_p^-(R_i) \approx -(1-\beta)R_i$ and finally $E[\tilde{R}_i^{\tau} - R_i^{(\tau)}] \leq (1-\beta)R_i$. Combining this with equation (4), it follows that the condition for the punishment (fine) $f_p(R_i^{(t)})$ of the participant i that acted as provider at round t becomes:

$$f_p(R_i^{(t)}) > \sum_{\tau=t+1}^{\infty} \frac{1}{N} \delta^{\tau-t}(-\Delta R_p^-) \cdot [a_i(b-v) + (1-a_i)pb]$$

$$\approx \frac{1}{N} \frac{\delta(1-\beta)}{1-\delta} R_i^{(t)} [a_i(b-v) + (1-a_i)pb] \stackrel{\triangle}{:=} \tilde{w_p}(R_i^{(t)}, a_i) \,. \quad (5)$$

Next, we consider that the fixed participant considered above has acted as a client j at the deviation round t, when transacting with a provider that succeeded in service provision. Reasoning similarly as in the proof for the case that the participant has acted as a provider in round t, we obtain:

$$f_c(R_j^{(t)}) > b(1-p) + \frac{1}{N} \frac{\delta}{1-\delta} R_j^{(t)} \frac{\frac{1-\beta}{N}(\frac{1}{\tilde{r}^{(t)}} - 1)}{1 + \frac{1-\beta}{N}(\frac{1}{\tilde{r}^{(t)}} - 1)} \quad (6)$$

$$\cdot [a_j(b-v) + (1-a_j)pb] \stackrel{\triangle}{:=} b(1-p) + w_c(R_j^{(t)}, a_j) \,.$$

■

Note that the lower bounds for the fines $f_p(R_i^{(t)})$ and $f_c(R_i^{(t)})$ in equations (5) and (6) provide the range for sufficiently large such punishments so that deviations from truthful reporting are unprofitable. These bounds depend on the probability of success in providing services of satisfactory quality a_i of the participant, which is private information. Recall that the probability of success is the hidden information that is revealed either by the Beta reputation metric or by the proposed one after many transactions. Therefore, a_i for a participant i can be approximated by $R_i \bar{r}$. On the other hand, $a_i(b-v) + (1-a_i)pb$ can be replaced by its upper bound $\max\{b-v, pb\}$, since $a_i \in [0,1]$.

Punishing both transacted parties upon disagreement in feedback reports is an approach that certainly punishes the lying party at the expense of an unfair punishment of the party that was sincere. This unfairness generates *social loss* that increases with the values of fines for disagreement. If fixed fines are employed, then their values should be high enough to satisfy that $f_p > \tilde{w_p}(R)$ and $f_c > b(1-p) + w_c(R)$ for the *maximum* values of $\tilde{w_p}(R_i^{(t)})$, $w_c(R_i^{(t)})$, which due to monotonicity arise when $R_i^{(t)}$ equals the highest possible rank, namely $1/\bar{r}$. On the contrary, when reputation-based punishments are employed, the exact values of $\tilde{w_p}(R)$ and $w_c(R)$ for each participant with rank R that acts as provider or client respectively are computed; thus, smaller and customized bounds for fines are then derived. We can quantify the social loss generated under fixed fines as follows: First, we calculate the mean reputation-based fine that is paid by an average provider or client assuming a certain distribution of the ranks in the market is known; the corresponding probability density function is denoted as $PDF(R)$. The ratio of the social cost of mean reputation-based fines over fixed ones for the provider (PSR) is given by the following formula:

$$PSR = \frac{\int_0^{\frac{1}{\bar{r}}} PDF(R) f_p(R) dR}{f_p(1/\bar{r})} < 1 \,. \quad (7)$$

A similar formula applies for the client (CSR). In the next section, we provide some related numerical results.

VI. Impact to Online Feedback Mechanisms

In this section, we discuss the application of our approach to existing online feedback mechanisms. Among the most important Internet auction environments are eBay, Yahoo! Auctions (http://auctions.yahoo.com), ePier (http://www.epier.com), and uBid (http://www.ubid.com). These online auction sites employ feedback mechanisms similar to that of eBay, to which we focus in the sequel. In fact, eBay which is a prominent example of an electronic market where a participant can act as both provider and client, which is one of the main features of our model. Besided this our model (of both the market and the feedback system) has has other important features in common with eBay, and thus it is appropriate for studying how submission of truthful feedback by the vast majority of participants can be enforced in eBay.

According to the feedback mechanism of eBay, clients and providers rate each other after their transaction, based on their satisfaction (i.e. utility) from the transaction. The feedback can

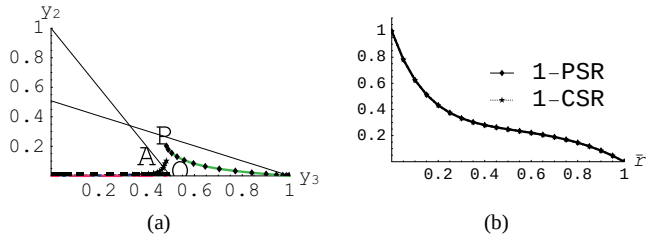
either be positive (+1), negative (-1), or neutral (0). Also, it is possible that no feedback is sent after a transaction. The feedback system of eBay publishes the number of positive ratings minus the number of negative ones (which is the reputation score) and the ratio of positive ratings over the total number of ratings for three different time periods: past month, past 6 months, past 12 months. In case of mutual negative feedback between transacted parties, a *feedback dispute* may be raised. During the dispute, if the provider accuses the client for not paying, then an *unpaid item notification* is sent to the client. If the notification for an unpaid item remains unanswered by the client and the provider issues an *unpaid item strike*, then the rating of the client is removed but its comment remains. During a feedback dispute, both transacted parties may decide to mutually withdraw their negative feedback to each other.

As already mentioned, eBay and its feedback scheme of eBay has important features in common with our model; namely: a) a participant can act as both provider and client, b) both transacted parties may submit feedback on their transaction and a feedback dispute (i.e. disagreement) may arise thereafter, and c) reputation is only taken into account for selecting providers. However, there are two apparent differences of our feedback scheme from that of eBay. We explain that these differences are actually insignificant. The first difference is that, in our model, the feedback by both transacted parties rates only the providing behavior. In eBay, the provider submits feedback on the behavior of the client, that is on her honesty for sending the payment or for not backing-off from auctions she won. This feedback is aggregated with the other ratings that this client previously received when acting either as provider or as client. Often a client initiating an auction can ban clients with low reputation from participating. If Paypal (http://www.paypal.com) or an analogous mechanism for securing money-transfers is employed, then fraud related to payments cannot actually arise. In such cases, submission of feedback for clients is unnecessary, as money transfers are guaranteed. Nevertheless, such feedback is still employed providers rate clients in order to deter strategic or malicious feedback by the latter, as in our scheme. Then, we can safely conclude that, with high probability, any disagreements (i.e. submission of negative ratings both for the client and for the provider) arise due to lying (either by the client or by the provider) on the performance of the provider. These coinside with the two cases of disagreement in our model! The second difference of our scheme with the eBay's one is the induction of fines upon disagreement in our case. In eBay, disagreement accounts for mutual negative feedback, while in our case it refers to different feedback ratings for the provider. Punishment for disagreement in eBay is done by counting the negative feedback for both transacting parties, and thus reducing their expected value in future transactions. This form of punishment depends on the current reputation values of the transacted parties, and it is actually an indirect implementation of our reputation-based fines! However, mutual withdrawal of negative votes in eBay allows avoiding this punishment and thus fraud may be left unpunished, as opposed to our mechanism. To summarize, our model is appropriate for analyzing eBay and its feedback scheme since the differences between them are minor.

Resnick *et al.* argue in [11] that there is low incentive for eBay participants to provide ratings. Users withhold negative ratings, because they are "nice" (e.g. in the hope of a reciprocal positive rating) or they fear retaliation. Thus, there is bias towards positive rating: only 0.6% of the ratings received by clients and 1.6% of the ratings received by providers are negative. Anonymity could be a solution against preferential or discriminatory rating. Often a user does not directly benefit by providing a rating, e.g. when the feedback is positive and users are competing. Resnick *et al.* found in [11] that only 51.7% of clients and only 60.6% of providers provide ratings in their transactions. Nevertheless, as argued in [11], the reputation system of eBay does provide the incentive for good behavior as its members belief that it works and act on the basis of this belief. Dellarocas and Wood suggest in [12] that clients should take into account missing feedback although it is very difficult for the average user to estimate this information. They estimated that, on the average, eBay clients walk away from a transaction while being satisfied 78.9% of the time, mildly dissatisfied 20.4% of the time and very dissatisfied 0.7% of the time. The corresponding estimates for providers are 85.7%, 13.7% and 0.6% respectively.

Next, we employ the aforementioned results of [11], [12] to estimate the population fractions of eBay that would submit feedback according to the three strategies (i.e. *True*, *Lie*, *Duck*) if our feedback scheme (with its disagreement fines) were in place. Providers and clients that do not submit any feedback can be safely considered to play *Duck*. As already mentioned, the corresponding population fractions are 48.3% for clients and 39.4% for providers. These fractions concern *average* estimates for providers and clients over *both* the success and the failure subgames. Next, we calculate the average fractions of providers and clients that play *Lie* in both subgames. We denote as p_{feed} the fraction of providers that submit feedback (consisting of p^+_{feed} positive and p^-_{feed} negative), p_{sat} the total fraction of satisfied providers, p^+_{sat}, p^-_{sat} the fractions of satisfied providers among those that submit positive and negative feedback respectively, and p^0_{sat} the fraction of satisfied providers among those that do play *Duck*. Then, it holds that $p_{sat} = p^+_{sat} \cdot p^+_{feed} + p^-_{sat} \cdot p^-_{feed} + p^0_{sat}(1 - p_{feed})$. We can safely assume that no satisfied providers submit negative feedback (i.e. $p^-_{sat} = 0$). Then, from [11], [12] and the fact that $p^0_{sat} \leq 1$, we obtain that $p^+_{sat} \geq 0.768$. Thus, we find that the pessimistic case for the fraction of providers that play *Lie* is 0.139. If we assume that $p^+_{sat} = 0.884$ (referred to as the average case), then the fraction of providers that play *Lie* becomes 0.069. Obviously, in the optimistic case (i.e. $p^+_{sat} = 1$), the fraction of liar providers becomes 0. Applying the above reasoning for clients, we obtain that their fraction that plays *Lie* is 0.202 in the pessimistic case, 0.101 in the average case and 0 in the optimistic case.

Now, we are able to evaluate the effectiveness of our feedback scheme in eBay. We set the price parameter b of



(a) (b)

Fig. 3. a) Evolution of strategies of eBay clients after successful service provision in the pessimistic (P), the average (A) and the optimistic (O) cases. b) $1 - PSR$ and $1 - CSR$ as functions of the mean reputation.

our model using the mean closing value of coin auctions, and specifically $b = 52.98$\$ [12]. This value of b has been calculated in [12] for a sample population of $N = 22287$, where 28% of them act as providers and 72% of them acting as clients. Therefore, we can assume that with probability q=0.28 the members of the population act as providers and with probability 1-q they act as clients. The mean reputation in this population is $\bar{r} = 0.838$, while reputation can be lognormally distributed according to Lim *et al.* [13]. Also, we assume that a utility for clients $u = 60$\$, a profit margin 10% (i.e. $v = 0.9 \cdot b$), a discount factor $\delta = 0.99$, $\beta = 0.6$ and $p = 1$. (Note that the choice $p = 1$ renders our model similar to the model of eBay transactions where the complete payments are transferred upfront.) We first consider the case of employing fixed monetary punishments according to Theorem 1. If we assume the pesimistic case for the percentage of liars in eBay, then fixed monetary punishments (regardless of their severity) would lead eBay to the all *Duck* stable equilibrium according to evolutionary dynamics 1, as depicted for clients in the success subgame in Figure 3(a). Nevertheless, should we decreased the population fraction that plays *Duck*, eBay would evolve to all *True* stable equilibrium. In particular, in the average and the optimistic cases, any fixed monetary fines f_p, f_c that satisfy $f_p \geq 0.13 > \tilde{w_p}$ and $f_c \geq 5 \cdot 10^{-5} > w_c$, lead eBay to all *True* stable equilibrium (see Figure 3(a)). Note that choosing $f_p \approx \tilde{w_p}$ and $f_c \approx w_c$ would result in very small basins of attraction for the [*True*, *True*] strategy pair. However, these basins of attraction expand with f_p and f_c at the expense of a greater social cost. This is intuitevely clear, because higher fines are more effective threats against *Lie* and *Duck*. We have experimentally proved that these results are also valid when reputation-based fines are employed, but we omit the details for brevity reasons. Note that in all three cases, if lower population fractions that play *Duck* were considered, then the market would always evolve to the all *True* equilibrium. This could be accomplished in eBay, if the submission of feedback for a transaction were mandatory within a timeout and enforced by e.g. doubling listing fees associated to the transactions in cases of ducking.

One could argue that the introduction of disagreement fines could deter participation in the online market. However, note that these fines are very small as compared to the price of a single transaction, as already demonstrated. Moreover, reputation-based fines are increasing functions of reputation. Since newcomers in the market have low reputation values, their fines are the lowest among the population. Also, the risk of a fraudulent transaction is present in eBay and other online markets. Rating negatively the provider for this fraudulent behavior is not easy at present in eBay (e.g. a feedback dispute is raised) and it can even be costly due to provider's retaliation. Therefore, a feedback scheme that employs disagreement fines can be more effective. Also, as explained, revenues due to the enforcement of fines are not expected to be significant in the long run due to the evolutionary stability of the truthful equilibrium. Nevertheless, fines could be collected by the market owner for the computational and communication overhead induced by the employment of this incentive mechanism.

Next, we evaluate the reduction of the social loss due to unfair punishments that is achieved when reputation-based fines are employed instead of fixed ones, which is expressed by $1 - PSR$ and $1 - CSR$ [see equation (7)]. As depicted in Figure 3(b), the social loss reduction per disagreement achieved by reputation-based fines is significant for providers and clients and decreases with mean reputation in the market.

As a future work, we plan to investigate other means that combined with fines would force the e-market to evolve to the desired equilibrium.

REFERENCES

[1] C. Dellarocas, *Reputation Mechanisms (ch. 13)*. Economics and Information Systems (T. Hendershott, ed.), Elsevier Publishing, December 2006. ISBN-10: 0-444-51771-5.

[2] T. G. Papaioannou, G. D. Stamoulis, "Reputation-based policies that provide the right incentives in peer-to-peer environments," *Computer Networks, Special Issue on Management in Peer-to-Peer Systems*, vol. 50, no. 4, pp. 563–578, March 2006.

[3] ——, "An incentives' mechanism promoting truthful feedback in peer-to-peer systems," in *Proc. of IEEE/ACM CCGRID (Workshop on Global P2P Computing)*, Cardiff, UK, May 2005.

[4] ——, "Enforcing truthful-rating equilibria in electronic marketplaces," in *Proc. of the IEEE ICDCS Workshop on Incentive-Based Computing*, Lisbon, Portugal, July 2006.

[5] C. Dellarocas, "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior," in *ACM Conference on Electronic Commerce*, October 2000.

[6] ——, "Goodwill hunting: An economically efficient online feedback mechanism for environments with variable product quality," in *Proc. of the Workshop on Agent-Mediated Electronic Commerce*, July 2002.

[7] N. Miller, P, Resnick, and R. Zeckhauser, "Eliciting informative feedback: The peer-prediction method," *Management Science*, vol. 51, no. 9, pp. 1359–1373, September 2005.

[8] R. Jurca, B. Faltings, "An incentive compatible reputation mechanism for the online hotel booking industry," in *Proc. of IEEE Conference on E-Commerce*, Newport Beach, SA, USA, May 2004.

[9] L. Samuelson, *Evolutionary Games and Equilibrium Selection*. The MIT Press, 1997. ISBN 0-262-19382-5.

[10] A. Jøsang, S. Hird, E. Faccer, "Simulating the effect of reputation systems on e-markets," in *Proc. of the 1st International Conference on Trust Management*, Crete, Greece, May 2003.

[11] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood, "The value of reputation on ebay: A controlled experiment," *Experimental Economics*, vol. 9, no. 2, pp. 79–101, June 2006.

[12] C. Dellarocas, and C. A. Wood, "The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias," in *Workshop on Information Systems and Economics (WISE)*, Irvine, CA, December 2005.

[13] Z. Lin, D. Li, B. Janamanchi, W. Huang, "Reputation distribution and consumer-to-consumer online auction market structure: An exploratory study," *Decision Support Systems*, vol. 41, no. 2, pp. 435–448, January 2006.