UNIL | Université de Lausanne
HEC Lausanne

IRM

INSTITUTE OF RESEARCH IN MANAGEMENT
INSTITUT DE RECHERCHE EN MANAGEMENT

# THE SOCIAL DYNAMICS OF ONLINE REPUTATION SYSTEMS

## WORKING PAPER 0606

**Anthea Monod**

**And**

**Jean-Claude Usunier**

# THE SOCIAL DYNAMICS OF ONLINE REPUTATION SYSTEMS

JEAN-CLAUDE USUNIER & ANTHEA MONOD

Abstract. This paper discusses the evolution of reputation information to recommendation information through online reputation systems. Firstly, a list of social dynamic terms that are relevant in the successful evolution of a reputation system is defined. Working from these definitions, the paper then proceeds to discuss mechanism design that leads to the successful implementation of a reputation system on the basis of a natural reputation setting, and in particular, a successful reputation system that transforms mere reputation information into recommendation information.

## 1. Introduction

In an interdisciplinary field that seeks to merge information technology and behavioral economics in order to achieve an understanding of efficiency, it becomes vital to make definitions clear and concise to have any hope of attaining a common goal. This paper sets out to succinctly define social dynamic terms that are relevant in the study of online reputation systems. Furthermore, in the context of the proposed definitions, the paper then proceeds to discuss the evolution of an online reputation system. In particular, it is argued that the best-case scenario is for the online reputation system to propagate recommendation information. Recommendation information emerges when reputation information advises the best course or choice and comes from a credible source; or, in other words, when reputation information is trustworthy. Working from the basis of these definitions, characteristics that an online reputation system should possess with this goal of recommendation information in mind are then discussed, and an evolutionary scheme in order to achieve this aim is proposed.

This paper is organized as follows: first, the evolution of online feedback to recommendation information is presented and the superiority of recommendation information over online feedback and even reputation information is discussed. Following this discussion, the paper aims to precisely define the terms that are relevant for achieving the goal of recommendation information. In particular, defining terms that are structural to the setting of online forums that a reputation system may take advantage of, as well as avoid, through appropriate design in order to attain recommendation information is a focal point. The idea of a

natural reputation setting is introduced, and social dynamics within such a setting are discussed. Using this lattice of social dynamics in a natural reputation setting as a basis for online reputation systems, features of mechanism design the evolutionary track of a reputation system from its emergence to its sustainability past a certain threshold point are presented, highlighting the possible downfalls that could lead to its collapse on this track.

## 2. Feedback, Reputation and Recommendation

In the computer science literature, reputation is often equated with trust and cooperation, and also on occasion with confidence (a supposed synonym of trust; Luhmann for a discussion on the differences between confidence and trust). In order to articulate these concepts and illustrate the reputation mechanism, consider the following example: an individual wishing to dine out may look at the reputation of various restaurants on an online reputation system (ORS). In this situation, the potential diner would be a trustor, the restaurants potential would be trustees, and reputation information provided by the online reputation system allows the potential diner to reduce risk. A consequence of trust as a decision-making factor is its implementation, that is, cooperation. If a particular restaurant, which initially had a good, disappoints the diner "trustworthy" reputation has failed in satisfying the diner and will receive, as a result, a bad report from this diner that will affect its previous good reputation.

In this particular case, dissatisfaction in the service encounter is the result of an interactive process where the trustor is victimized. Cooperation is possible without economic trust, especially in cases where there is a high power differential between the parties, and one party is obliged to cooperate even at the risk of being exploited for the mere sake of survival (e.g. when a small industrial subcontractor accepts orders from a large manufacturing company).


Simply put, online feedback is basic information posted on the Internet by users that provides reflective cues of feedback targets; feedback targets may be products, services, institutions, or even other users. It may be automated (e.g. Google) or manually posted by a user and

take on various forms, including ratings (choice of position on an evaluative scale for various characteristics), rankings, votes (choice of preferred candidate among various alternatives), or qualitative comments. Implicit or explicit, there is always an element of value judgment in feedback information (good, bad, neutral), particularly in the case of manually posted feedback. Unless the source of feedback has a previously established and outstanding expertise, and is fully impartial, the individual feedback information must be averaged across all reporters in some way to increase its usefulness and credibility.

2.1. **Reputation, Trust and Cooperation.** Once feedback is aggregated, it becomes reputation information; the process of aggregation adds a dimension of value over online feedback, since it provides raw data summary statistics of the raw data of online feedback. Reputation information is a more convenient form of data, since the summary statistics serve as an acceptable substitute over reading each and every individual posting. It is also likely to be more accurate, especially for those users who are seeking information in order to decide whether or not to engage in a transaction but, because of time constraints, can only read a portion of the online postings available. It is also more accurate for those users who are simply unable to remember the content of each posting. This is particularly true as the number of online feedback postings grows. The issue of aggregating feedback proves to be nontrivial, particularly in decentralized networks, since various problems arise in locating and retrieving the relevant information in the network, assessing the credibility of reporters and/or reports, and transforming individual data into a reputation score in a reliable manner and at a reasonable cost. Furthermore, the data transformation process must be scalable in order to maintain its effectiveness under a growing user database. A reputation system is designed in order to address these problems. In general, a reputation system "helps people decide whom to trust, encourages trustworthy behavior, and deters participation by those who are unskilled or dishonest through collecting, distributing, and aggregating feedback about the participants' past behavior." [21].

Reputation information can then inspire trust among users. This is achieved via two complementary mechanisms: reputation information reduces the risk of potential trustors by providing them with the means to make an informed decision; simultaneously, it is a means for the potential trustees to send a signal and thus provides them with a means to increase their chances of economic exchange, as they are now able to increase others' willingness to trade with them [3].

Inspired trust leads to cooperation among users. The trust mechanism that emerges from reputation information has a positive effect on cooperation (i.e. if the reputation information proves to be positive, this will encourage cooperation; conversely, if the reputation information is negative, this will undermine cooperation). This discourages trustees from misbehaving overall: because potential trustors will use their reputation information as a factor in decision-making, it is in the trustees' interest to exert a higher level of effort in transactions.

2.2. **Definitions of Trust.** In game theory and economics, there exist various interpretations of trust depending upon the circumstance. Strategically, trust can be defined in a two-person strategic setting. According to Berg et al., a two-stage and two-player interaction reflects a situation of trust if the following conditions are met (see Berg et al. 1995, [2]):

(i) One of the players, the trustor, is first asked whether she gives a second player, the trustee, the right to take a decision which affects the payoff of both players,
(ii) By trusting, the trustor is put in a position of risk,
(iii) Honoring trust benefits the trustor at a cost to the trustee, and finally,
(iv) Placing and honoring trust Pareto-dominates the subgame perfect equilibrium outcome, for which the trustor refrains from placing trust.

For this paper's purposes, it is sufficient that the trustor is put in a position of risk by trusting, or that honoring trust benefits the trustor at a cost to the trustee, to assume that trust exists within the system. As long as one of these two conditions are met, it will be assumed that trust is present, regardless of other circumstances.

Trust can also be called upon as a notion, rather than a key definition, that serves as an explanation for certain phenomena. For Luhmann [20], trust can be seen as a means of reducing complexity. In the rational decision-making approach, agents make the decision to trust drawing from logical deduction and based on reason; however, according to Luhmann, trust may be called upon if rational decision making fails (that is, when agents act in a sense that would contradict reasonable deduction). This type of interpretation is further supported by propositions of Ortmann et al. (2000, [23]) and Cox et al. (2002-2004, [5], [6], [7], [8], [9]), who argue that choosing to trust can result from confusion, lack of saliency, curiosity or framing effects.

2.3. **Personalization and Trust.** Similarity can be assessed from being acquainted with a person or by identifying shared characteristics with the person without necessarily having had prior contact. Characteristics-based trust (Zucker, 1986, [28]) increases trustees' trustworthiness. Because similarity fosters predictability of behavior, it decreases vulnerability (i.e. reduces risk perceived by the trustor), and therefore increases the propensity to trust a particular target. An extreme form of similarity-based trust is interacting only within one's ingroup members (Greif, 1994, [16]), a choice predicated on the basis of an extended similarity assumption: "if they are like me, they will behave as I will"; as a consequence, behavior (seemingly) becomes fully predictable. If reputation information is associated with general features describing the source such as age and gender, it is transformed into recommendation.

Trust, and the implied cooperation that exists as a result of trust, are consequences of recommendation. A reputation system provides recommendation information when it becomes personalized to a certain extent. The superiority of recommendation information over reputation information lies in the improved focus of reputation information that is more closely aligned to the user's own characteristics and preferences. Under heterogeneous attitudes and preferences, recommendation information reduces information asymmetries more effectively, thereby increasing overall market efficiency.

The personalization of a reputation system can be broken down as follows:

- Full personalization is the case when providers of feedback are completely defined; that is, non-anonymous, with sufficiently declared characteristics, attitudes, and preferences in order to generate a complete online profile.
- Implied personalization allows the user to select reputation information from a set of anonymous reporters on the basis of their common socio-demographic characteristics (age, gender, income, ethnic background, etc.), attitudes, self-reported behavior, or self-stated preferences. Under implied personalization, users do not have a formal identity or profile.

## 3. Structural Properties of Online Reputation Systems and Social Dynamics

The use of information technology for commerce and interaction in general is subject to certain social dynamics; it is also conditioned by human motivation, and constrained by cognitive abilities (in cases that deal with complex systems). In order to fulfill their purpose, online reputation systems must first and foremost motivate the report

of individual feedback, and in particular, they must motivate users to report their true opinions, (Bolton, 2002, [4]). The following section outlines the key elements of social dynamics that must be taken into account in order for this preliminary aim to be attained, and their relation to the establishment of trust within the system.

As emphasized by Dellarocas et al. (2004, [11]), "the success of online reputation systems depends on the sustained voluntary contribution of feedback by community members." It is tempting for people to be only users of reputation information contributed by others without oneself taking the time and effort to report. If everybody free-rides on the online reputation system, consulting it without leaving feedback, this would imply by the logic of backward induction that the online reputation system would not exist to begin with. For this not to be the case, the users must display some altruistic orientation, or at least some goodwill. Other motivations to report such as direct reciprocity (posting positive/negative information to reward or punish a target) or indirect reciprocity (explained hereafter) make pure free riding an issue mostly related to the minimum degree of participation needed for the online reputation system to be effective. Extensive free riding (i.e. low rate of reporting) is acceptable provided that the sample of reporters is large enough and representative of the base population of users.

3.1. **Altruism and Goodwill.** Altruism is a form of behavior corresponding to unconditional cooperation and kindness, without any ulterior motives (Jenks, 1990, [17]). Put differently, as by Andreoni in [1], altruism is a behavioral premise; altruism, by definition, does not arise as the result of having experienced altruistic behavior bestowed by another. In particular, altruistic acts are ones that allot a positive payoff to another agent at a cost to the performing agent (according to Trivers [25]). In technical terms defined by Fehr and Fischbacher in [15], an altruistic agent takes the payoffs of some other reference agent (who may himself be altruistic or not) as a basis for his own actions. Altruistic agents, therefore, always act in order to maximize the payoffs of the reference agent. Trust is easily established in a system where altruistic users dominate; however, feedback and reputation may not be credible as there is a strong bias towards overwhelmingly positive reports.

Goodwill can be seen as a weaker form of altruism, and synonymous with the understanding of impure altruism, as proposed by Andreoni, also in [1]. The main difference in the understandings of altruism and goodwill is that there is the element that the agent's own utility is taken into consideration when she chooses to exhibit goodwill, through

the increased utility reaped from the convictin of being a good citizen (i.e. experiencing a "warm glow"), and perhaps also from having others believing the same. Agents exhibiting goodwill also experience an increase in their own utility when acting to increase the utility of others. However, they do not take the payoffs of some other agent as a basis for their actions. Rather, they act in such a way so that the utility of others increases, with the ulterior motive that their own utility will be increased when the others view them as good citizens, and will avoid acting in such a way that would bring them disutility when others believe them to be bad citizens. Concretely, when willingness to cooperate is considered, altruism is the stronger notion and so implies goodwill. Conversely, agents who do not cooperate when cooperation would damage their personal image as good citizens does not necessarily lack goodwill, but is definitely not altruistic. Trust also emerges easily where agents exhibit goodwill; trust is also sustainable for as long as goodwill is present within the society.

3.2. **Reciprocity.** Indirect reciprocity refers to general cooperative behavior within an already-established society, with an anticipated reciprocal payoff of more than just the proverbial "warm glow" in return. Drawing from the formulation of Nowak and Sigmund in [22], indirect reciprocity can take on two forms:

(i) Upstream reciprocity: First A helps B, and then B helps C
(ii) Downstream reciprocity: First A helps B, and then C helps A

Extending the idea of indirect reciprocity beyond biological notions, this paper proposes a third form of indirect reciprocity through implied reciprocity. This occurs when A first helps B, then C, having observed A help B, also helps B. This notion is particularly relevant in economic exchange, as it is precisely this mechanism that generates reputation information among users, which then implies trust and cooperation as previously discussed above. When C decides to help B, C has received an implicit signal that B must be trustworthy, as B has been able to induce A to help B. Thus being trustworthy can be viewed implicitly as B helping A. B receives an indirect reciprocal payoff when C helps B. The following example illustrates this mechanism: B owns a shop, where A shops regularly; by buying goods from B's shop, A helps B. C sees that A shops at B's shop, and thinks that B must provide good service and sell high-quality goods, which is why A shops there. So then C decides to also buy goods from B. The reciprocity comes into play when B implicitly helps A by providing good service and selling high-quality goods. B receives a reciprocal payoff when C buys goods from B.

The key understanding behind the concept of indirect reciprocity is that the cooperative behavior occurs within a society. The reciprocal payoff associated with indirect reciprocity does not come from the agent with whom one cooperates, but either from another agent within that society (in the case of downstream reciprocity), or from the general benefit that comes from the upkeep and maintenance of a public good (as implied by upstream reciprocity). This demonstrates that this social dynamic entails elements of goodwill or impure altruism.

"Help" can be understood in various ways in regard to the emergence and sustainability of trust. In the case where "helping" corresponds to merely posting (truthful) feedback, all three forms of indirect reciprocity (upstream, downstream and implied) reduce to the maintenance of a public good (namely, the reputation system) and thus can imply the existence of goodwill.

When "helping" means posting only positive feedback, upstream reciprocal behavior can be seen as goodwill: first A reports positively on B, B then behaves as a good citizen and also reports positively on C. This is also true in the case of downstream reciprocal behavior: first A reports positively on B, displaying good citizenship; C, having seen A's behavior also decides to exhibit good citizenship and reports positively on A. In these two cases, when "helping" means only reporting positively, trust may emerge, but may not necessarily be sustainable. In the case of implied reciprocity, when A reports positively on B, C may also report positively on B because C may consider A to be in a position of authority in order to report positively and decides to also report following A's authority. This situation also demonstrates goodwill and good citizenship, because A either reports truthfully from the position of authority to do so, or reports positively because of some either pure or impure altruistic orientation; in either case, A is exhibiting goodwill. C, by also reporting positively following A's actions, is demonstrating trust in A, a characteristic of good citizenship. B may or may not exhibit goodwill; if A reports positively as a truthful report, B indeed exhibits goodwill, but if A reports positively because A is (purely or impurely) altruistic, B need not necessarily exhibit goodwill. Thus, it is entirely possible for B to also be a goodwill agent under this scenario. When all three agents exhibit goodwill and report positively, trust emerges easily, but may not be sustainable if the number of positive-only reports dominates the total number of reports. Another interpretation may be that A reports positively on B because B has established a good reputation and lives up to it, so that when A reports positively, A is actually reporting truthfully. Then, by virtue of B's good reputation, C also reports truthfully and positively. This is a particular case where posting only positive feedback and posting

truthful feedback are equivalent; it can be considered a situation where all agents exhibit goodwill by displaying good citizenship, since both A and C are reporting truthfully and B lives up to a good reputation. Again, trust emerges easily in this situation and moreover is sustainable in this case of truthful reporting. Trust also emerges easily in the case when A and C report positively because they display some altruistic orientation, but may not be sustainable if the number of positive-only reports is overwhelming.

"Helping" could also be understood as posting negative feedback on one user in order to distort his or her reputation, thereby artificially inflating a third party's reputation. This understanding of "help" describes negative indirect reciprocity, where the action taken by the user appears to be defective behavior on the surface; it cannot be described by any of the three previously mentioned cases of indirect reciprocity.

In settings where reputation systems exist to aggregate feedback in an online market forum, "helping" can have a less subtle, more commercial understanding. "Helping" can simply mean buying from the reputation target. In this sense, neither upstream nor downstream reciprocity describe any reciprocal behavior at all, they merely describe normal trade relationships in any market. Reputation information can still emerge from normal trade behavior. After repeated transactions within the community, a sense of brand image may emerge if many agents buy from one particular agent or institution and with it, a sense of reputation. Reputation may emerge faster and easier when such normal trade relationships are coupled with the idea that "help" can also mean cooperating by providing better service; in this case, implied reciprocity is the key mechanism that effectively propagates trust within the reputation system.

Direct reciprocity is similar to, but simpler than, indirect reciprocity. Direct reciprocity refers to cooperative behavior in a two-player setting where cooperation is the strategy employed with an expected return from the other player. In other words, A helps B, then B helps A in return. As in the case of indirect reciprocity, the understanding of "helping" that is most conducive to inspiring trust is the simultaneous, commercial understanding of buyer and seller cooperating in their respective capacities.

3.3. **Collusion and Retaliation.** Collusion is the act of collaborating with other agents toward the aim of undermining a rival. In the case of complete, active collusion, all agents within the online community

are partitioned into an ingroup of colluders and an outgroup of non-colluders. All ingroup agents are colluders and consciously collaborate with each other with the common aim of undermining one, several, or all outgroup agents (the rivals).

Another case of collusion that can exist is that of passive collusion which is similar to the idea of to negative indirect reciprocity. In the case of negative indirect reciprocity, there is one single active member, A, who reports negatively on agent B in order to augment agent C's reputation. But then C, whether he is willing and conscious of it or not, has been involved in the collusive act performed by A, even though he has not taken any action. In this situation, C is a passive colluder.

While the existence of all forms of collusion in a reputation system leads to the collapse of trust on the general level when both ingroup and outgroup members are taken into account, trust still exists within the system, since a premise for active collusion is trust amongst the collaborating agents. In the context of collusion, we see that trust is not necessarily virtuous. In considering a system where there exists both ingroup and outgroup members, the ingroup members' informational advantage over outgroup members introduces a bias into the system, which precisely leads to market inefficiencies caused by asymmetric information. However, even in the case of this vicious social dynamic, a market-efficient, stable equilibrium can still be achieved when the system comprises only ingroup members (i.e. the system is robust to the entry of new agents). In this situation, trust emerges and is sustainable within the system, even though the conditions in which it exists may not be desirable. Two caveats exist to this supposed equilibrium: first, this understanding of trust may emerge as a response to panic (following the formulation of Luhmann et al.), and second, trust may be an enforcement mechanism (cooperation for fear of banishment). In the latter case, the situation at hand is not one of trust according to the assumed definition, but rather one of anti-ingroup trust: here, not trusting (and thus not cooperating with) other ingroup members is the strategy that would present a risk for the agent (namely, the risk of being outlawed); trust is basically forced upon all members. While systematically, the collusive equilibrium with trust may be equivalent to other equilibriums demonstrating trust, this is not the case from a social point of view, in particular for a reputation system.

Retaliation can be thought of as negative direct reciprocity, where a negative initial report results in a retaliatory response that is also negative, thereby introducing bias into the system. In the initial phase of reputation formation, if reports are positive and neutral, trust may still emerge, even if retaliatory intentions are present among users. It

is only once negative feedback starts being reported that retaliation becomes active, and its destructive effects evident. When retaliation begins to dominate within a reputation system, the only stable equilibrium exists is the "equilibrium of terror", where no one reports at all for fear of having his own reputation spoiled. While this outcome solves the problem of biased reporting, it also undermines trust within the social network.

"Anarchy", is a social dynamic that has the exact opposite sentiments of pure altruism. All users are destructive to the point of irrationality, as they are so committed to a lose-lose outcome that they act without any consideration of their own losses. It can be considered as altruism's polar case. There exists one fundamental technical symmetry between "anarchy" and altruism: the fact that an agent's own utility is never taken into account under both altruistic and anarchic behavior. "Anarchy" can also be more precisely named as negative altruism, as altruism by definition involves only the utility of a reference agent (or entity, when the altruistic agent aims to maximize the global utility of all users within a system) is the only factor taken into account for the altruistic agent's decision to act. This is also true in the case of negative altruism (Fehr and Gächter, 1998, [14]); negative altruistic users consider only the utility of the reference entity (that is, the global society of users), with the aim in mind of its collapse. Even under negative altruism, trust and cooperation may still emerge when it is necessary for survival, but will nevertheless be difficult to sustain. While negative altruism may exist in some users in an online community, generalized negative altruism is unlikely to emerge on a full scale since the existence of such a generalized sentiment would inhibit the emergence of the online community and reputation system. Therefore the emergence and sustainability of trust may still be robust to this social dynamic.

## 4. Natural Reputation Systems

In order to discuss the design of a reputation system, a natural setting will first be presented where the previously described social dynamics coexist and the emergence of reputation in such a setting. Under this setting, we will then discuss the effects of these social dynamics on an online reputation system.

We define a natural reputation system to be the reputation system that emerges naturally from any given structure that constitutes the social dynamics of any given human grouping. For instance, a natural reputation system exists in any firm in any industry, where the employees on all hierarchical levels (managers, subordinate employees,

maintenance staff alike) are all users in the reputation system. A particular example of a setting for a natural reputation system is also a school or university, where faculty, administrative staff, and students are all users in the reputation system.

4.1. **A Folk Theory of Reporting.** Reporting entails issuing a judgment statement on others; one who reports is behaving as a "person qualified to comment critically" according to the Collins dictionary. Most religions as well as folk wisdom advise against judging others, as there is an element of arrogance related to considering oneself as "qualified" in order to make critical statements. It may also be the case that the person who reports may not actually be "qualified" to make an informed judgment; she may lack technical or contextual knowledge. Another practical reason for advising against judgment is that reporting could be motivated by reasons unrelated to the actual feedback target. When judgment is critical, and the target of feedback is people or institutions, it may negatively affect their (or its) interests. Often the ramifications of judgment are so severe that a need for legal regulation arises; both slander (defamation in a transient form, such as by spoken words or gestures) and libel (publication of defamatory matter in a permanent form, such as printed material) are governed by laws. As a consequence, there is an underlying social understanding that one should not judge others, or at least refrain from explicitly expressing their judgment.

The folk rule of thumb would be of the nature "If you don't want to be judged, don't judge others". In some societies, the avoidance of judging can extend as far as complete repression of consumer complaints, even when there are legitimate reasons for expressing dissatisfaction. In other societies, the social stigma of judging others is taken advantage of and can be used by potential targets to coax positive feedback from others: for example, in German restaurants, waiters usually ask diners after each course if they enjoyed the dish. In this situation, if the food was not to their liking, diners often find it difficult to express dissatisfaction with the meal right in front of those who have prepared it, so that they automatically answer "yes", or to tone down their dissatisfaction to a mere mumble.

When it comes to judging, it is easier to do so in the absence of those being judged, or behind their backs. It is also easier to judge when the target of judgment is not likely to identify the source of the judgment; that is, when the judgment is anonymous. Double blind reporting is a precautionary mechanism that may encourage truthful reporting and ease the process of judgment for the reporters, but cannot always be

used as there naturally exist asymmetries of anonymity in some situations: for example, students may evaluate a professor and a course anonymously, but professors always rate (grade) students nominally. Such asymmetries of anonymity often arise when there is an imbalance of power between reporters; asymmetry is usually granted to the less powerful group. In the folk theory of judging others, there is a certain reluctance, which can be partly, but not fully, offset by granting anonymity to raters.

4.2. **Avoidance of Reporting.** Reporting is costly in both a deterministic and stochastic manner: there is a fixed cost of spending time and effort in preparing and posting reports, as well as an uncertain cost arising from potential retaliation from targets as a result of reporting. Avoiding being a source of reputation information is the best way to avoid being a target of reputation information. Reciprocity will generally act against the emergence and the sustainability of reputation systems. As a consequence, most individuals have no reputation. By default, they are considered as having some level of reliability and performance as the average score of individuals (or organizations) for which (reliable) reputation information is available. The lack of a reputation is an attractive characteristic for the large majority of social actors who do not want to be obliged to proactively manage their social image.

Targets of reputation information are also likely to avoid promoting the use of reputation systems; most companies are not motivated to create reputation systems for their own products and services. For instance, some users or companies who play a more focal role in the community cannot avoid being a target of reputation. In general, they prefer to send messages that affect their reputation themselves, rather than having others (such as critics, consumers, media, etc.) send such messages. Great precautionary measures are taken to avoid potential negative consequences of third party-based reputation information on their image. For this reason, companies invest heavily in advertising and brand image development; with these tools, they are able to unilaterally create their own discourse on their products and services with a large margin for exaggeration and deviation from their actual quality. Spending on brand image development may also include creating an image of good citizenship for the company, perhaps due to the social stigma of judging and the arrogance associated with rating themselves positively. For this reason, some companies will invest heavily in charities or other proceeds that will generate this image of goodwill; for example, Philip Morris, the cigarette manufacturer, goes

as far as largely broadcasting the health dangers associated with smoking and heavily investing in programs aimed at helping smokers to quit.

Reputation systems may help to create a brand image for these companies at a substantially lower cost than advertising (as low as 1% of the cost of advertising), however, there is still a stochastic element as to the nature of the brand imagethat actually emerges from the reputation system. Providing a medium where users who need not be affiliated with the company nor feel a personal responsibility to promote the company's products and services can be difficult to control, with a chance that the brand image that results is negative. As such, providing a reputation system for its own products is a difficult strategy that is rarely employed by companies.

Companies that are self-confident in their product and services are more capable of enjoying the low costs associated with using a reputation system to increase their brand image. An important factor in corporate self-confidence is the nature of the product and its targeted demographic group. For instance, Lush, a large multinational British company that manufactures cosmetics for women with an emphasis on all-natural ingredients and handmade techniques, provides an online forum for international unsers to chat with each other about the products. In the case of Lush, the demographic group of users is overwhelmingly female, and a high proportion is environmentally conscious. As such, their positive brand-image is likely to be sustainable, as their company and product characteristics are largely robust to attackers. They are a sufficiently large company, who target females (females tend to be brand-loyal, especially when cosmetics are concerned), and appeal to the environmentally conscious community by exhibiting an image of good citizenship by producing earth-friendly products. Companies with a lower level of self-confidence in their product and services are better off choosing a high-cost, but controlled, way of disseminating reputation information.

4.3. **Other Regarding Behavior and Stability within Natural Reputation Systems.** As noted above, a key property of such natural reputation settings is that posting reputation information incurs a cost to the user, reducing the motivation to report. For this reason, altruism (pure altruism) and goodwill (impure altruism) are *prime facie* positive drivers of reputation in such a natural reputation setting. However, there exist limitations to these drivers with regard to the sustainability of a reputation system: an excess of both pure and impure altruism in users may result in overwhelmingly positive feedback reports (since positive feedback is expected from pure and impure altruists, as both types of users by definition cooperate, almost unconditionally). This

in turn leads to the degeneration of the significance and effectiveness of feedback and reputation information. Thus, there must exist a proportion of pure and impure altruists within such a natural reputation setting that will stimulate and sustain useful and substantial feedback and reputation information. (See Appendix A for a formalized argument of this notion.)

Furthermore, this maximum must also take into account the existence of free-riders in the system. Altruism, together with the incurred cost of reporting, triggers the response of free-riding. When there are sufficiently many altruists who cooperate unconditionally, free-riders who observe this behavior are able to take advantage of the availability of reputation information posted by these altruistic users without reporting themselves (they do not contribute to the maintenance of reputation information), which inhibits the emergence of the natural reputation system. For this additional reason, the proportion of altruistic users within the system must be non-negative so that the online reputation system may emerge, and finite. (See Appendix A.)

4.4. **Decentralization and Multilateralism as Properties of Natural Reptuation Systems.** In natural reputation settings as described above, there is no central mediator whose role is to manage all users' reputation and reputation information in general within the system (i.e. the central mediator does not retrieve feedback, collate and aggregate this feedback into reputation information, and disseminate the information). As such, in order to be applicable to the general level of reputation settings, any online reputation system must be decentralized in some way so that no central authority manipulates information.

Another key property of natural reputation settings that must be considered is the multilateral nature of reporting among users within the setting: all users within a natural reputation system are free to report feedback on any given user to any other user/s. Users may also solicit information on their own reputation from other users in order to react reciprocally. In the one-shot, unilateral case, reciprocity does not exist, since by definition, for a given pair of users, only one user may report on the other. In the multi-period game, if the same player reports on the other in every period, reciprocity does not exist. If each of the two players sequentially reports on the other, reciprocity exists. In the bilateral case, both users in a given pair of users may report on each other. They may do so simultaneously, or sequentially. In the one-shot bilateral and multilateral game, reciprocity only becomes significant in the case of sequential reporting. In the multi-period bilateral and multilateral game, reciprocity is significant in both simultaneous and

sequential reporting.

In practice, in natural reputation settings, reciprocity between users hinders the emergence of reputation information altogether. With neutral reciprocity, users mutually agree to not report on each other, resulting in no feedback and therefore no reputation information. With negative reciprocity, users who have received an initial negative report retaliate with a negative report, which eventually leads to an equilibrium of terror (as described previously, which here can be seen as equivalent to neutral reciprocity), where again, no reputation information emerges at all. With positive reciprocity, a user will respond with a positive report in exchange for an initial positive report; numerous occurrences of this phenomenon results in skewed and therefore insignificant (objective) reputation information, as in the case of excess altruism within the system. Of these three cases, neutral reciprocity is the defensive strategy and the one that is practically most employed in such natural reputation settings, impeding the emergence of reputation information.

With indirect reciprocity, also a multilateral case of reputation propagation, a fundamental concern is that the settings are personalized to an extent for each user, so that the user's own indirect reciprocal payoff is sufficient motivation for her to continue acting in such a manner. Again, neutral (indirect) reciprocity is the defensive strategy, so that indirect reciprocity is found to be a weak driver of reputation information. Finally, the property of multilateral reporting, together with the decentralized characteristic, implies a lack of resilience to collusion and in- and outgroup orientation within natural reputation settings.

Table 1 (see Appendix B) provides a summary of the effects of the previously discussed social dynamics on the emergence (startability) and sustainability of natural reputation systems.

4.5. **Reputation Information within Natural Reputation Settings.** In a natural reputation setting, reputation information is more likely to be higher demanded when there are:

- high information asymmetries;
- behavior and/or outcomes of behavior that cannot easily be observed and verified (lack of transparency is a general characteristic of social systems);
- disadvantaged users are aware of the information asymmetry, and the unequal distribution of gains that results;
- the asymmetry of information cannot be overcome by signaling;
- reputation information becomes a fundamental decision factor for exchange.

## 5. Desirable Properties for an Online Reputation System

This paper will now examine a number of properties of online reputation systems which are relevant for mechanism design. It is our view that decentralization is a desirable property of an online reputation system (based on P2P networks or any kind of distributed information systems architecture). This is not the case of actual online reputation systems where a central actor collects, aggregates, and publishes reputation information. Also, unilateral online reputation systems where raters are not rated themselves would be favorable. In most real-world situations, for instance between a service customer and a service provider, there is asymmetry in size, role, numbers, power, etc. so that ratees are not raters and vice-versa. An argument that a free, unilateral, semi-anonymous online reputation system is preferable in that it reduces direct reciprocal behavior.

5.1. **Anonymous vs. Non-anonymous.** This paper will now examine a number of properties of online reputation systems which are relevant for mechanism design. It is our view that decentralization is a desirable property of an online reputation system (based on P2P networks or any kind of distributed information systems architecture). This is not the case of actual online reputation systems where a central actor collects, aggregates, and publishes reputation information. Also, unilateral online reputation systems where raters are not rated themselves would be favorable. In most real-world situations, for instance between a service customer and a service provider, there is asymmetry in size, role, numbers, power, etc. so that ratees are not raters and vice-versa. An argument that a free, unilateral, semi-anonymous online reputation system is preferable in that it reduces direct reciprocal behavior.

5.2. **Anonymous vs. Non-anonymous.** This paper will now examine a number of properties of online reputation systems which are relevant for mechanism design. It is our view that decentralization is a desirable property of an online reputation system (based on P2P networks or any kind of distributed information systems architecture). This is not the case of actual online reputation systems where a central actor collects, aggregates, and publishes reputation information. Also, unilateral online reputation systems where raters are not rated themselves would be favorable. In most real-world situations, for instance between a service customer and a service provider, there is asymmetry in size, role, numbers, power, etc. so that ratees are not raters and vice-versa. An argument that a free, unilateral, semi-anonymous online reputation system is preferable in that it reduces direct reciprocal behavior.

5.3. **Anonymous vs. Non-anonymous.** In a fully non-anonymous online reputation system, rater identity is fully known. Because their identity is fully known (it is additionally assumed that this identity can be verified), raters are likely to be targets of retaliation in case of negative reporting, especially when reputation information is in the form of qualitative reviews, which contain statements that may be considered as libel. If in addition the cost of reporting and the deep-seated reluctance to judge others described above is considered, fully non-anonymous online reputation systems will face the obstacle of reporting sources being scared of and therefore unwilling to post reputation information. Sometimes, pseudonyms are used to try and palliate source shyness. However, pseudonyms must be stable over time: if a source constantly changes pseudonyms, it is even worse than anonymity in the sense that it is potential indication of untruthful reporting (liars cannot be caught, but untruthful information may sometimes be identified). Full anonymity at the other extreme, although making it easier for sources to report, also facilitates irresponsible, manipulative reporting and collusion (colluders cannot be identified since they are anonymous).

A semi-anonymous online reputation system, with implied personalization and stable/reliable pseudonyms, has many advantages over both a fully anonymous and a fully non-anonymous online reputation system. A semi-anonymous online reputation system divides the rater profile into two sets of information: a pseudonym and a set of source characteristics (sociodemographics, preferences, attitudes, level of expertise concerning the target object being rated). It allows users of reputation information to check whether the source corresponds to them, that is, to assess source credibility in their perspective. Furthermore, the online reputation software may allow users to compute reputation scores on the sole basis of individual sources that correspond to particular characteristics: a woman may be willing to obtain reputation scores only from other women, or only from other women in a certain age bracket, etc. This feature corresponds to the passage from reputation to recommendation without fully losing anonymity, with non-anonymity being a key deterrent of motivation to report.

5.4. **Public vs. Private.** An online reputation system may be open to everyone (public) or may be restricted to a limited number of users that are allowed to use it privately. To illustrate this, let us imagine an online reputation system for Business-to-Business (B2B) products or services where access to the online reputation system is restricted to a limited number of companies which are customers or suppliers of certain items. This creates a small community where reputation information is difficult to spread in a natural, objective form. Not only would reporters try to manipulate information (e.g. customers report below

true performance in order to artificially obtain improved performance from suppliers), but collusion would be easier and more tempting, with some targets inducing sources to negatively report on a particular competitor in order to lower its market share or eventually drive it out of the market. A small and closed world is on average more susceptible to the "defeat of anonymity": reporters are theoretically anonymous but almost everybody knows - or believes they know - who stands behind a particular report. A closed online reputation system may work as a fantasized ingroup where an exacerbation of reciprocal behavior, both positive and negative, is likely to occur. A small world is also less likely to reach the critical mass of reporters needed for meaningfully aggregating reputation scores. This may be aggravated since raters tend to be more motivated to report the negatives (dissatisfaction, complaints, incidents) than the positives (satisfactory experiences, performance beyond expectation). As a rule, humans tend to consider it normal when things work and to consider it as unacceptable when their expectations are not met.

An open, public online reputation system is preferable because it allows any source with some experience and/or expertise with the product or service in question to post information. Such sources should naturally make it clear in which role they were, what was the context of their experience and level of expertise. Consider the case of an online reputation system that allows airline passengers to report about flight delays. Airlines are not the sole responsible for flight delays: airport authorities or maintenance services may be accountable as well as aircraft manufacturers, tour operators, possibly other passengers or weather conditions that have been ignored by passengers because they are not explained by the crew. In such complex online reputation systems, it is important that each actor is allowed to express her views while being clearly identified as belonging to a particular class of actors.

5.5. **Free vs. Paid Online Reputation System.** There are some arguments in favor of a paid online reputation system (raters and/or users of recommendation information pay to access the online reputation system): it allows an online reputation system manager to collect resources in order to manage the platform which will collect information, check it, and aggregate it into reputation scores. Micropayments can also be used when source is paid for reporting truthfully, creating an incentive compatible mechanism that discourages untruthful reporting (Jurca & Faltings, 2005, [18]).

However, this paper argues that online reputation systems should remain free of charge in terms of access cost and without direct reward

for those contributing information. Online reputation systems are typical public goods: sources post information on the basis of goodwill, not financial compensation. If their reporting action has some financial motivation, it is only indirect and far-flung: their hope is in an improvement in product or service performance and/or a decrease in cost; they expect these improvements for themselves as well as for others. The motivation is, however, relatively tenuous and fixing a summplementary access cost in addition to the costs described above is likely to discourage sources from reporting and users from accessing reputation information.

5.6. **Metric.** There are two basic sorts of reports: either sources rate targets with numbers on a scale for various aspects of the object (rating) or they post information in the form of a detailed, written report in free form (review). Both forms, quantitative and qualitative metrics, may be present simultaneously in an online reputation service. This is the case for many online reputation systems dedicated to restaurants and food services. In the case of reviews, aggregation is difficult, if not impossible. It may be performed only by the user who consults the online reputation system, and personally processes the reviews consulted. This solution makes sense when there are relatively small numbers of reporters for each item being evaluated (e.g., as above, a restaurant or any kind of service outlet in a local community).

Rating scales provide the user of reputation information with more operative scores. Since the information is entered on numerical scales, a score can be computed across reporters for a particular item, however less easily across both reporters and items since items may be of different importance or even incommensurable. Consider the example of an online medical reputation systems where patients rate doctors. There are two main approaches to an online reputation service in terms of what kind of experience patients report:

- Positive-only ("best doc" approach) where patients only report positive experiences [1]
- Negative-only or negative-also reputation systems ("complaint" approach) which allow patients to file reports that include a clear negative message on the doctor's performance with the risk that it may be unfair, manipulative, or systematic (the same patient will rate all doctors negatively)

A number of different metrics (rating scales) may be used. In the real world different aspects of the patient doctor interaction may be evaluated. The underlying metric of positive and negative online feedback

[1] A concrete example of positive only reputation system can be found on the Swiss (centralized) consumer comparison website Comparis for health insurance companies (caisses d'assurance maladie, at comparis.ch).

can be made (at least numerically) comparable by using first-order differences rather than absolute values.

The positive-only scenario has advantages over a metric with negatives. The reason for not envisioning an online reputation system based on complaints, especially if reputation is managed centrally and raters' identities are fully known, is that the survival chances of such a system are petty given the risk of legal suits for libel (i.e. a form of negative reciprocal action). Considering vested interests of reputation targets against the online reputation system that potentially attack their public image, backward induction would lead potential promoters of such negative-centralized online repuation systems to renounce their attempts to promote it quite early in the stage of project feasibility analysis.

The extent of the scale has also to be looked at. A scale with two few positions, such as [-1; 0; +1] in eBay does not offer room for nuance. A '-1' appears as a severe punishment and is consequently quite rare. A positive-only rating scale with five to seven positions (1-5 or 1-7), excluding zero which appears in colloquial language as marked by a negative value judgment ("he is a zero"), appears as the most recommendable scale.

A last issue in the metric side of online reputation systems is the additional feature of default reporting. Default reporting is about assigning an arbitrary value (e.g. the mean value of the rating scale) to some or all the subjects who were supposed to report but did not. It can be used to compensate for the silent majority that does not report because they are either free-riders (willing to consult the online reputation system but not to contribute information) or reluctant to judge. Default reporting may also counterbalance the potential asymmetry between motivation to report satisfaction and motivation to report dissatisfaction. Imagine an airline that collects reputation information on its flights but has only 30% of its passengers that contribute by answering to the on-board satisfaction survey. Default reporting is analogous to replacing missing values by the mean of documented observations in a database. There is one fundamental drawback: if there is a large number of default reports versus few true reports, the reputation score is likely to converge to the middle of the scale and thus becomes meaningless.

5.7. **Reputation vs. Recommendation.** As argued above, reputation information becomes more credible for users when they can assess to what extent the sources of reports share among them a number of characteristics, a process that fosters similarity-based trust. Such a

process transforms reputation into recommendation.

## 6. Measures of the Efficiency of Online Reputation Systems

### 6.1. **The Traps of Low and High Regime Reputation Systems.**

The low regime takes place for an online reputation system when either too few users report and the system does not produce reliable information, or raters do not believe that their feedback will be taken seriously by targets, or the gap between their expectations and the average perceived performance of targets is too small. In this sense, if the standard deviation of quality of service (QoS) and/or price across reputation targets is minimal, it is meaningless to report. In markets where QoS and/or price are homogeneous, online reputation systems make less sense. Due to backward induction, it becomes meaningless for reputation sources to comparatively observe the rating of targets and to contribute to the public good by rating targets. If a reputation system does not really take off, it will slowly peter out.

High regime occurs when many users report. The high regime is likely to occur in the contrary circumstances as described above, especially when

1. the gap between expectations and average performance is high;
2. users believe that their feedback is taken seriously by the targets; and
3. there is significant variance in ratees' QoS and/or prices.

The users' belief that their feedback is taken seriously by the reputation targets will be activated by observing that the targets react to aggregate ratings. In the high regime, motivation to report is conditioned by a significant level of variance in information reported. In the case of no variance in reputation information, the high regime is not likely to maintain itself. In a sense, it is a self-fulfilling prophecy that, in the long run, the success of online reputation systems, necessarily leads to their decline. This is much more likely to be the case in a closed community, as that envisioned in B2B example above, than in an open community where some people exit the online reputation system but also new users constantly join loyal and regular raters. Constantly having new raters and new users join the online reputation system community is a requisite for sustainability.

Since most reputation settings are asymmetric in nature (oligopolies with few suppliers and many customers), both the low regime and the high regime are threats for the sustainability of an online reputation system. Under low regime conditions, an online reputation system

might never really take off, then gradually decline, and finally disappear. Online reputation systems are especially threatened by collapse during their start-up period, as targets tend not to adjust in the short run and motivation to report is largely related to gaps between expectation and actual performance (Parasuraman et al., 1985, [24]), variance of performance across competing targets, and the actual observation that ratings lead targets to increase effort.

In the high regime, an online reputation system may become a threat for the less efficient suppliers: they may lose market share against better-rated competitors. It is important to stress that virtually everyone in the target population is liable to lose from an online reputation system: since less efficient suppliers will be obliged to change their level of effort from low to high, the higher performance suppliers will progressively lose their comparative advantage. Being few and powerful, targets are likely to create a coalition and insist that the online reputation system be dismantled. A further argument in favor of decentralization of online reputation system is that it is more difficult to dismantle a decentralized, anonymous online reputation system under the pressure of targets.

6.2. **Online Reputation System Process and Outcome Measures.** Favorable outcomes of online reputation systems only occur when the reporting process works smoothly. For this reason, measures of process efficacy/efficiency are also needed, especially at the launch of an online reputation system. The easiest way to assess the efficacy of an online reputation system is to look at whether it succeeds in starting from scratch, surviving, and then growing over time. If there are not sufficiently many reporters, reputation information will not be reliable and users will not take into account for their decisions to trust products, people, or services. An early collapse is all the more likely when reports do not accumulate quickly enough to reach the critical mass that makes reputation information reliable. In order to be successful, an online reputation system must first reach a threshold participation rate for both reporters and users of reputation information. Reporters must report frequently enough so that they update their feedback and obsolete information does not distort aggregate reputation scores, and they must report truthfully (please see Table 2 below). These are classified as individual, process-based measures of efficiency: they are related to a favorable cost-benefit analysis for sources of reports.

If the process runs smoothly, the first indicator of outcome is individual satisfaction of both reporters and users within the online reputation

system. It then can be measured through surveys (providing a measure of reputation for the reputation system as a whole). However, there are also indirect effects at the individual level. They are listed in Table 2 (see Appendix B) as individual, outcome-based measures of online reputation system efficiency. The online reputation system may be consulted but not actually used for decision-making. Also, it may be consulted, but with consequences that agents evaluate ex post as more or less consequential for their own welfare. Finally, a significant fraction of agents may consider that their ability to improve individual decision making through online reputation systems is limited because the images of reality provided by the reputation system are at odds with their individual experience. This is more likely to be true when preferences are heterogeneous and not taken into account. Once again, this illustrates a fundamental superiority of a recommendation system which significantly reduces the probability that reputation information is irrelevant to users' own personal profile and preferences, and therefore to their decision-making. Another individual outcome-based measure of efficiency at the individual level is whether people perceive reputation information as useful and whether they actually use reputation information to make trusting decisions, which may possibly result in favorable economic consequences for them.

The two quadrants of process-based and outcome-based aggregate measures of efficiency of an online reputation system in Table 2 have been the most studied, especially on the basis of eBay data. Aggregate measures of process efficiency deal with the size and overall behavior of the user base. Measures of online reputation system efficiency in terms of aggregate outcomes are economic variables listed in Table 2, relating to price and quality levels, market structure and efficiency in resource allocation.

## 7. Conclusion

Most elements of social dynamics described in this paper are detrimental either to the emergence or to the sustainability of an online reputation system. Pure altruism and direct reciprocity (neutral, positive and negative) undermine online reputation systems by encouraging either non-reporting or biased reporting that reduce the credibility of reputation information. Forms of indirect reciprocity may help the propagation of reputation information (sustainability) but they have no effect on the emergence of online reputation systems. Furthermore, upstream reciprocity involves the risk of a courtesy bias beyond the start-up phase while implied reciprocity implies the risk of mimicking behavior (B rates C in a certain way because she observes that A did the same previously). Among the few social dynamics that should be used to promote an online reputation system and make it an objective

tool for reputation assessment is mostly goodwill. However, an online reputation system is not sustainable on the sole basis of goodwill. As a consequence, it appears that the design of an online reputation system should principally bypass the potentially negative consequences of social dynamics on the reporting of information and the credibility of such information. In order to achieve this goal, this paper argues in favor of a free, public, decentralized, semi-anonymous (i.e. with implied personalization), unilateral, recommendation-oriented online reputation system, with non-negative ratings.

## Appendix A

Claim: Following Section 5.4, there exists a strictly positive, finite proportion of pure and impure altruists within a natural reputation setting that will stimulate and sustain useful and substantial feedback and reputation information. Moreover, this level of altruism in the system generates an extremum, and is a maximum in particular.

*Sketch of proof:* If the number of users is sufficiently large, then a sufficiently smooth function $f$ may be constructed, describing the significance and effectiveness of feedback and reputation information as a function of the proportion of (both pure and impure) altruistic users in the system. Furthermore, it is reasonable to assume that effectiveness is nonnegative in altruism.

When there exists no altruism at all in the setting, the effectiveness is zero. Similarly, when there exists too much altruism, the effectiveness is again zero (as argued above). Thus, by Rolle's Theorem, we have a stable equilibrium of effectiveness at some nonzero proportion between these two points. Moreover, this stable equilibrium is an extremum, and by appropriate specification of the function, it is a maximum.

Remark 1: This maximum need not be unique, and need not be the global maximum.

Remark 2: Since altruism triggers the response of free-riding, we may also specify the function to take into account the free-riding effect.

## Appendix B: Tables

Table 1: Summary of Effects of Social Dynamics on Startability and Sustainability

| **Behavioral Property** | **Emergence/Startability** | **Sustainability (Resistance to Collapse)** |
|---|---|---|
| Altruism (Pure Altruism) | + | − : Excess altruism results in futile reputation information |
| Goodwill (Impure Altruism) | + : Limited by Cost Benefit Analysis | − : Goodwill disintegrates in the long run, the system is not sustainable on goodwill alone |
| Direct Positive Reciprocity | − or at best 0 (in the case of Neutral Reciprocity) | − or at best 0 : This form of reciprocity may lead to a mild form of collusion in the long run, and encourages a bias towards positive reputation information (courtesy bias), as in eBay |
| Direct Negative Reciprocity | − : Disappointed users are more likely to report than satisfied users, leading to a starting negative bias, which may then result in a reputation system that exists to signal the bad users instead of providing objective reputation information on the whole population | − : Risk of stabilizing at the equilibrium of terror |
| Indirect Upstream Reciprocity | 0 | +/0 : Encourages "helping" and cooperative behavior, but may also indirectly propagate a courtesy bias |
| Indirect Downstream Reciprocity | Unlikely to exist in natural and online reputation systems | 0, in general : May be sustainable in systems where reporting is at least bilateral |
| Collusion | Meaningless with regard to the startability of a reputation system, unless colluders are those who initiated the online reputation system | − : Results in biased positive or negative reputation information and leads to either insignificance of reputation information, or even newcomers who are threatened or blackmailed into cooperating by colluders in extreme negative cases |
| Ingroup vs. Outgroup Orientation | −/+ : Ingroup orientation is detrimental to the startability of the reputation system; outgroup orientation is favorable | +/0 : Reputation system is sustainable when all users of a reputation system are outgroup members (i.e. there exists no ingroup orientation); may also be sustainable even in cases where ingroup orientation exists, provided the proportion of ingroup orientation in the system is sufficiently low |

Table 2: Measures of the Efficiency of an Online Reputation System in terms of Process and Outcome Variables

| | Individual Agent | Aggregate |
|---|---|---|
| Process | Cost-benefit analysis | Number of users/growth of user-base over time |
| | Participation Level 1 (reporter): posting feedback | Quality of Reputation information (truthful? credible? useful?) |
| | Participation Level 2 (user): looking at reputation information | Collusion among groups of users |
| | Frequency of use | Reputation of ORS itself |
| | Truthful reporting | |
| Outcome | Satisfaction with ORS | Price decrease |
| | Use of reputation information in decision-making | Market structure |
| | Perceived ability to improve individual decision-making through ORS | Quality of items |
| | Consequences for agent of using ORS on own welfare | Market efficiency in general |

## References


[1] J. Andreoni *Giving with impure altruism: applications to charity and Ricardian equivalence*, The Journal of Political Economy **97(6)** 1447–1458 (1989).

[2] J. Berg, J. Dickaut, K. McCabe *Trust, reciprocity and social history*, Games and Economic Behavior **10** 122–142 (1995).

[3] G. Bolton, E. Katok, A. Ockenfels *Bridging the trust gap in electronic markets*, Discussion Papers on Strategic Interaction, Strategic Interaction Group, Max Planck Institute of Economics **2002-26** (2002).

[4] G. Bolton, E. Katok, A. Ockenfels *How effective are online reputation mechanisms?*, Discussion Papers on Strategic Strategic Interaction, Strategic Interaction Group, Max Planck Institute of Economics **2002-25** (2002).

[5] J. Cox *On the economics of reciprocity*, Mimeo, University of Arizona (2001).

[6] J. Cox *Trust and reciprocity: implications of game triads and social contexts*, Mimeo, University of Arizona (2003).

[7] J. Cox *How to identify trust and reciprocity*, Games and Economic Behavior **46** 260–281 (2004).

[8] J. Cox, D. Friedman *A tractable model of reciprocity and fairness*, Mimeo, University of Arizona (2002).

[9] J. Cox, K. Sadiraj, V. Sadiraj *Trust, fear, reciprocity and altruism*, Mimeo, University of Arizona (2002).

[10] C. Dellarocas *The digitization of word-of-mouth: promise and challenges of online feedback systems*, Management Science **49(10)** 1407–1424 (2003).

[11] C. Dellarocas, M. Fan, C. Wood *Self-interest, reciprocity, and participation in online reputation systems*, MIT Sloan School of Management working paper 4500-04 (2004).

[12] C. Dellarocas, P. Resnick *Online reputation mechanisms: a roadmap for future research*, Summary report of the First Interdisciplinary Symposium on Online Reputation Mechanisms (2003).

[13] Z. Despotović *Building trust-aware P2P systems: from trust and reputation management to decentralized e-commerce applications*, EPFL PhD thesis 3313 (2005).

[14] E. Fehr, S. Gächter, *Reciprocity and economics: the economic implications of homo reciprocans*, European Economic Review **42(3–5)** 845–859 (May 1998).

[15] E. Fehr, U. Fischbacher *Why social preferences matter – the impact of non-selfish motives on competition, cooperation and incentives*, The Economic Journal **112** C1–C33 (2002).

[16] A. Greif *Cultural beliefs and the organization of society: a historical and theoretical reflection on collectivist and individualist societies*, Journal of Political Economy **102(5)** 912–950 (1994).

[17] C. Jenks *Varieties of altruism*, Mansbridge, Jane J. (ed.) Beyond Self-Interest, University of Chicago Press, 53–70 (1990).

[18] R. Jurca, B. Faltings *Reputation-based service level agreements for web services*, EPFL preprint (2005).

[19] R. Jurca, B. Faltings, L. H. Vũ, J.-C. Usunier, K. Aberer *Reputation mechanisms for web services*, EPFL preprint (2005).

[20] N. Luhmann *Familiarity, confidence and trust: problems and alternatives*, Gambetta, Diego (ed.) Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford, Blackwell, chapter 6, 94–107 (2000).

[21] P. Resnick, R. Zeckhauser, E. Friedman, K. Kuwabara *Reputation Systems*, Communications of the ACM **43(12)** 45–48 (2000).

[22] M. A. Nowak, K. Sigmund *Evolution of indirect reciprocity*, Nature **437** 1291–1298 (2005).
[23] A. Ortmann, J. Fitzgerald, C. Boeing *Trust, reciprocity and social history: a re-examination*, Experimental Economics **3** 81–100 (2000).
[24] A. Parasuraman, V. Zeithaml, L. Berry *A conceptual model of service quality and its implications for future research*, Journal of Marketing **49(4)** (1985).
[25] R. Trivers *Social evolution*, Benjamin-Cummings Publishing Company, Menlo Park, California (1985).
[26] J.-C. Usunier, M. Brülhart *Verified trust: reciprocity, altruism and randomness in trust games*, CEPR Discussion Paper **4578** (2006).
[27] L. H. Vũ, M. Hauswirth, K. Aberer *QoS-based service selection and ranking with trust and reputation management*, preprint (2005).
[28] L. Zucker *Production of trust: institutional sources of economic structure, 1840-1920*, Staw, B. M. and Cummings, L. L. (eds.) Research in Organizational Behavior, JAI Press, Greenwich, Connecticut, volume 8, 53–111 (1986).

Université de Lausanne, École des Hautes Études Commerciales
*E-mail address*: Jean-Claude.Usunier@unil.ch, Anthea.Monod@unil.ch