THE DIGITIZATION OF WORD-OF-MOUTH: PROMISE AND CHALLENGES OF ONLINE REPUTATION MECHANISMS

Chrysanthos Dellarocas

Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139
dell@mit.edu

Current Draft: October 1, 2002

ABSTRACT:

Online reputation mechanisms harness the bi-directional communication capabilities of the Internet in order to engineer large-scale word-of-mouth networks. They are emerging as a promising alternative to more established assurance mechanisms, such as branding and formal contracting, in a variety of settings ranging from online marketplaces to Internet search engines. At the same time, they are transforming a concept that had traditionally fallen within the realm of the social sciences into an engineering design problem. This paper surveys our progress in understanding the new possibilities and challenges that these mechanisms represent. It discusses some important dimensions in which Internet-based reputation mechanisms differ from traditional word-of-mouth networks and surveys the most important issues related to designing, evaluating and using them. It provides an overview of relevant work in game theory and economics on the topic or reputation. It further discusses how this body of work is being extended and combined with insights from computer science, information systems, management science and psychology in order to take into consideration the special properties of online mechanisms such as their unprecedented scalability, the ability to precisely design the type of feedback information they solicit and distribute, and challenges associated with the volatility of identities and the absence of many familiar contextual cues in online environments.

1 INTRODUCTION

A fundamental aspect in which the Internet differs from previous technologies for mass communication is its bi-directional nature: Not only has it bestowed upon organizations a low-cost channel through which to reach audiences of unprecedented scale but also, for the first time in human history, it has enabled individuals to almost costlessly make their personal thoughts and opinions accessible to the global community of Internet users.

An intriguing family of electronic intermediaries are beginning to harness this unique property, redefining and adding new significance to one of the most ancient mechanisms in the history of human society: *online reputation mechanisms,* also known as *reputation systems* (Resnick, Zeckhauser, Friedman and Kuwabara, 2000) are using the Internet's bi-directional communication capabilities in order to artificially engineer large-scale word-of-mouth networks in online environments.

Online reputation mechanisms allow members of a community to submit their opinions regarding other members of that community. Submitted feedback is analyzed, aggregated with feedback posted by other members and made publicly available to the community in the form of member *feedback profiles.* Several examples of such mechanisms can already be found in a number of diverse online communities (Table 1).

Perhaps the best-known application of online reputation mechanisms to date has been as a technology for building trust in electronic markets. This has been motivated by the fact that many traditional trust-building mechanisms, such as state-enforced contractual guarantees and repeated interaction, tend to be less effective in large-scale online environments (Kollock 1999). Successful online marketplaces, such as eBay, are characterized by large numbers of small players, physically located around the world and often known to each other only via easily changeable pseudonyms. Contractual guarantees are usually difficult or too costly to enforce due to the global scope of the market and the volatility of identities. Furthermore, the huge number of players makes repeated interaction between the same set of players less probable, thus reducing the incentives for players to cooperate on the basis of hoping to develop a profitable relationship.

Web Site	Category	Summary of reputation mechanism	Format of solicited feedback	Format of feedback profiles
eBay	Online auction house	Buyers and sellers rate one another following transactions	Positive, negative or neutral rating plus short comment; ratee may post a response	Sums of positive, negative and neutral ratings received during past 6 months (see Section 3)
eLance	Professional services marketplace	Contractors rate their satisfaction with subcontractors	Numerical rating from 1-5 plus comment; ratee may post a response	Average of ratings received during past 6 months
Epinions	Online opinions forum	Users write reviews about products/services; other members rate the usefulness of reviews	Users rate multiple aspects of reviewed items from 1-5; readers rate reviews as “useful”, “not useful”, etc.	Averages of item ratings; % of readers who found a review “useful”
Google	Search engine	Search results are rank ordered based on how many sites contain links that point to them (Brin and Page, 1998)	How many links point to a page, how many links point to the pointing page, etc. (see Section 5.6)	No explicit reputation profiles are published; rank ordering acts as an implicit indicator of reputation
Slashdot	Online discussion board	Postings are prioritized or filtered according to the ratings they receive from readers	Readers rate posted comments	

Table 1 Some examples of online reputation mechanisms used in commercial websites.

Online reputation mechanisms have emerged as a viable mechanism for inducing cooperation among strangers in such settings by ensuring that the behavior of a player towards any other player becomes publicly known and may therefore affect the behavior of the entire community towards that player in the future. Knowing this, players have an incentive to behave well towards each other, even if their relationship is a one-time deal. As I discuss in Section 3, a growing body of empirical evidence seems to demonstrate that these systems have managed to provide remarkable stability in otherwise very risky trading environments.

The application of reputation mechanisms in online marketplaces is particularly interesting because many of these marketplaces would probably not have come into existence without them. It is, however, by no means the only possible application domain of such systems. Internet-based feedback mechanisms are appearing in a surprising variety of settings: For example, Epinions.com encourages Internet users to rate practically any kind of brick-and-mortar business, such as airlines, telephone companies, resorts, etc. Moviefone.com solicits and displays user feedback on new movies alongside professional reviews and Citysearch.com does the same for restaurants, bars and performances. Even news sites, perhaps the best embodiment of the unidirectional mass media of the previous century, are now encouraging readers to provide feedback on world events alongside professionally written news articles.

The proliferation of online reputation mechanisms is already changing people's behavior in subtle but important ways. Anecdotal evidence suggests that people now increasingly rely on opinions posted on such systems in order to make a variety of decisions ranging from what movie to watch to what stocks to invest on. Only five years ago the same people would primarily base those decisions on advertisements or professional advice. It might well be that the ability to solicit, aggregate and publish mass feedback will influence the social dynamics of the 21st century in a similarly powerful way in which the ability to mass broadcast affected our societies in the 20th century.

The rising importance of online reputation systems not only invites, but also necessitates rigorous research on their functioning and consequences. How do such mechanisms affect the behavior of participants in the communities where they are introduced? Do they induce socially beneficial outcomes? To what extent can their operators and participants manipulate them? How can communities protect themselves from such potential abuse? What mechanism designs work best in what settings? Under what circumstances can these mechanisms become viable substitutes (or complements) of more established institutions, such as contracts, legal guarantees and professional reviews? This is just a small subset of questions that invite exciting and valuable research.

This paper surveys our progress so far in understanding the new possibilities and challenges that these mechanisms represent. Section 2 discusses some important dimensions in which Internet-based reputation mechanisms differ from traditional word-of-mouth networks. The discussion clarifies why this otherwise ancient concept merits new study. Section 3 presents an overview of eBay's feedback mechanism, perhaps the best-known online reputation mechanism to date. It summarizes initial field evidence on the mechanism's properties and formulates the most important questions relating to designing, evaluating and using such mechanisms. The following two sections survey our progress in developing a systematic discipline that can help answer those questions. First, Section 4 provides an overview of relevant past work in game theory and economics. Section 5 then discusses how this body of work is being extended in order to take into consideration the special properties of online mechanisms. Finally, Section 6 summarizes the main points of the paper and lists opportunities for future research.

2 ONLINE REPUTATION MECHANISMS: AN ANCIENT CONCEPT IN A MODERN SETTING

Word-of-mouth networks constitute an ancient solution to a timeless problem of social organization: the elicitation of good conduct in communities of self-interested individuals who have short-term incentives to cheat one another. The power of such networks to induce cooperation without the need for costly and inefficient enforcement institutions has historically been the basis of their appeal. Before the establishment of formal law and centralized systems of contract enforcement backed by the sovereign power of a state, most ancient and medieval communities relied on word-of-mouth as the primary enabler of economic and social activity (Benson, 1989; Greif, 1993; Milgrom, North and Weingast, 1990). Many aspects of social and economic life still do so today (Klein, 1997).

What makes online reputation mechanisms different from word-of-mouth networks of the past is the combination of (a) their unprecedented scale, achieved through the exploitation of the Internet's low-cost, bi-directional communication capabilities, (b) the ability of their designers to precisely control and monitor their operation through the introduction of automated feedback mediators and (c) new challenges introduced by the unique properties of online interaction, such as the volatile nature of online identities and the almost complete absence of contextual cues that would facilitate the interpretation of what is, essentially, subjective information.

Scale enables new applications. Scale is essential to the effectiveness of word-of-mouth networks. In an online marketplace, for example, sellers care about buyer feedback primarily to the extent that they believe that it might affect their future profits; this can only happen if feedback is provided by a sufficient number of current customers and communicated to a significant portion of future prospects. Theory predicts that a minimum scale is required before reputation mechanisms have any effect on the behavior of rational agents (Bakos and Dellarocas, 2002). Whereas traditional word-of-mouth networks tend to deteriorate with scale, Internet-based reputation mechanisms can accumulate, store and flawlessly summarize unlimited amounts of information at very low cost. The vastly increased scale of Internet-based reputation mechanisms might therefore make such mechanisms effective social control institutions in settings where word-of-mouth previously had a very weak effect. The social, economic and perhaps even political consequences of such a trend deserve careful study.

Information technology enables systematic design. Online word-of-mouth networks are artificially induced through explicitly designed information systems (feedback mediators). Feedback mediators specify who can participate, what type of information is solicited from participants, how it is aggregated and what type of information is made available to them about other community members. They enable mechanism designers to exercise precise control over a number of parameters that are very difficult or impossible to influence in brick-and-mortar settings. For example, feedback mediators can replace detailed feedback histories with a wide variety of summary statistics, they can apply filtering algorithms to eliminate outlier or suspect ratings, they can weight ratings according to some measure of the rater's trustworthiness, etc. Such degree of control can impact the resulting social outcomes in non-trivial ways (see Sections 5.2-5.4). Understanding the full space of design possibilities and the consequences of specific design choices introduced by these new systems is an important research challenge that requires collaboration between traditionally distinct disciplines, such as computer science, economics and psychology, in order to be properly addressed.

Online interaction introduces new challenges. The disembodied nature of online environments introduces several challenges related to the interpretation and use of online feedback. Some of these challenges have their roots at the subjective nature of feedback information. Brick-and-mortar settings usually provide a wealth of contextual cues that assist in the proper interpretation of opinions and gossip (such as the fact that we know the person who acts as the source of that information or can infer something about her through her clothes, facial expression, etc.). Most of these cues are absent from online settings. Readers of online feedback are thus faced with the task of making sense out of opinions of complete strangers. Other challenges have their root at the ease with which online identities can be changed. This opens the door to various forms of strategic manipulation. For example, community members can build a good reputation, milk it by cheating other members and then disappear and reappear under a new online identity and a clean record (Friedman and Resnick, 2001). They can use fake online identities to post dishonest feedback for the purpose of inflating their reputation or tarnishing that of their competitors (Dellarocas, 2000; Mayzlin, 2002). Finally, the mediated nature of online reputation mechanisms raises questions related to the trustworthiness of their operators. An important prerequisite for the widespread acceptance of online reputation mechanisms as legitimate trust building institutions is, therefore, a better understanding of how such systems can be compromised as well as the development of adequate defenses.

3 A CASE STUDY: EBAY'S FEEDBACK MECHANISM

This section presents an overview of eBay's feedback mechanism, perhaps the best-known online reputation mechanism to date. It summarizes initial field evidence on the mechanism's properties and motivates the need for a systematic discipline of online reputation mechanism design and evaluation.

Overview of eBay's feedback mechanism

eBay is the world's largest online marketplace. Founded in September 1995, eBay is the leading online marketplace for the sale of goods and services by a diverse community of individuals and businesses. Today, the eBay community includes 49.7 million registered users, and is the most popular shopping site on the Internet when measured by total user minutes[1].

On eBay, users are known to each other through online pseudonyms (eBay IDs). When a new user registers to the system, the only information that eBay verifies, is that her email address is valid[2]. Since there are many ways to sign up for anonymous free email accounts, this system means that anyone who wants to remain anonymous has the option to do so.

Most items on eBay are sold through English auctions[3]. A typical eBay transaction begins with the seller listing an item he has on sale, providing an item description (containing text and, optionally, photos), a starting bid, an optional reserve price and an auction closing date/time. Buyers then place bids for the item up until the auction closing time. The highest bidder wins the auction. The winning bidder sends payment to the seller. Finally, the seller sends the item to the winning bidder.

It is easy to see that the above mechanism incurs significant risks for the buyer. Sellers can exploit the underlying information asymmetries to their advantage by *misrepresenting an item's*

[1] Source: MediaMetrix, 2001.
[2] This is true for buyers only. To register as a seller, eBay also requires a valid credit card for verification purposes.
[3] eBay also supports Dutch auctions but these are rarely used.

true attributes (e.g. lying about its true quality) or by *failing to complete the transaction* (i.e. keep the buyer's money without sending anything back.)

It is obvious that without an adequate solution to the above adverse selection and moral hazard problems, sellers have an incentive to always cheat and/or misrepresent the attributes of their items. Expecting this, buyers would either not use eBay at all or place very low bids that would lead to a "market for lemons" (Akerlof, 1970). To address these problems, eBay uses online feedback as its primary trust building mechanism[4]. More specifically, following completion of a transaction, both the seller and the buyer are encouraged to rate one another. A rating can be a designation of the transaction as "positive", "negative" or "neutral", together with a short text comment.

eBay aggregates all ratings posted for a member into that member's *feedback profile*. An eBay feedback profile consists of four components (Figures 1 and 2):

A. A member's *overall profile makeup*: a listing of the sum of positive, neutral and negative ratings received during that member's entire participation history with eBay.

B. A member's *summary feedback score* equal to the sum of positive ratings received by unique users minus the number of negative ratings received by unique users during that member's entire participation history with eBay.

C. A prominently displayed "*eBay ID Card*", which displays the sum of positive, negative and neutral ratings received during the most recent six month period (further subdivided into ratings received during the past week, month and past six months.)

D. The *complete ratings history*, listing each individual rating and associated comment posted for a member in reverse chronological order.

[4] In addition to its feedback mechanism, eBay offers its members the option of using escrow services at extra cost. However, so far the percentage of transactions that opt for the use of those services is very low.

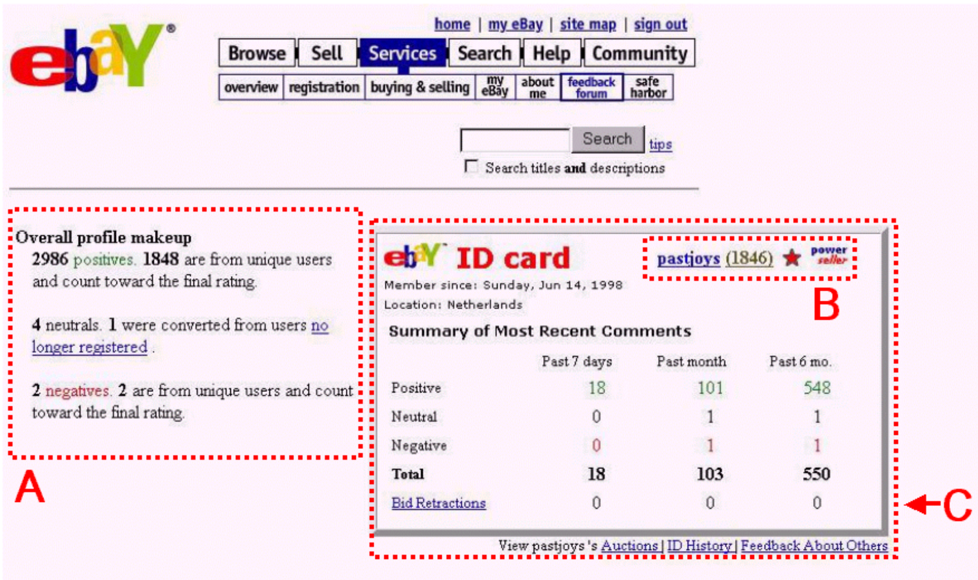

Figure 1 eBay member profile summary

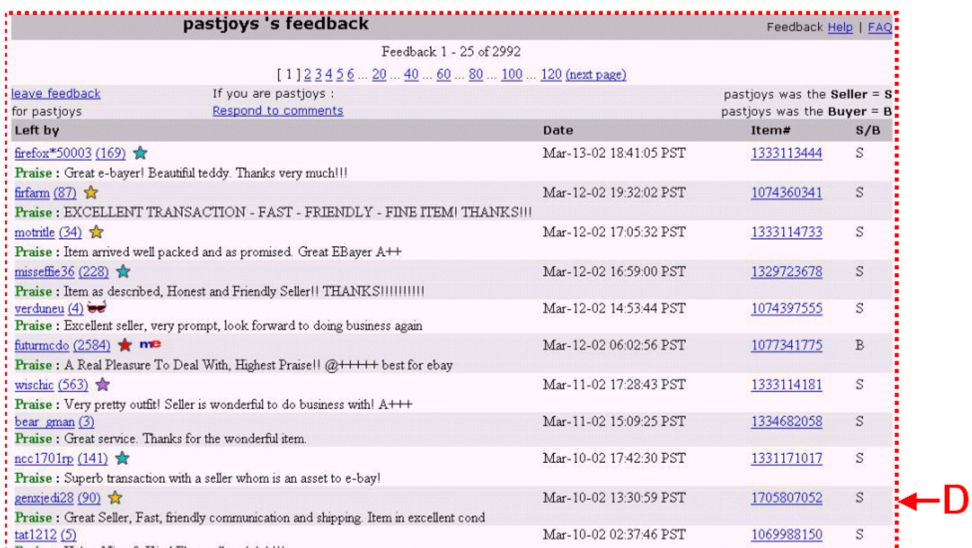

Figure 2 Detailed feedback history

Seller feedback profiles are easily accessible from within the description page of any item for sale. More specifically, all item descriptions prominently display the seller's eBay ID, followed by his summary feedback score (component B in Figure 1). By clicking on the summary feedback score, prospective buyers can access the seller's full feedback profile (components A, B and C) and can then scroll through the seller's detailed ratings history (component D).

eBay's feedback mechanism is based on two important assumptions. The first assumption is that members will indeed leave feedback for each other. Feedback is currently voluntary and there are no concrete rewards or penalties for providing it (or failing to do so). The second assumption is that, in addition to an item's description, buyers will consult a seller's feedback profile before deciding whether to bid on a seller's auction. Based on the feedback profile information, buyers will form an assessment of the seller's likelihood to be honest in completing the transaction, as well as in accurately describing the item's attributes. This assessment will help determine whether they will indeed proceed with bidding, as well as the bid amount. Sellers with "bad" profiles (many negative ratings) are therefore expected to receive lower bids or no bids to their auctions. Knowing this, sellers with long horizons will find it optimal to behave honestly even towards one-time buyers in order to not jeopardize their future earnings on eBay. At equilibrium, therefore, the expectation is that buyers will trust sellers with "good" profiles to behave honestly and sellers will indeed honor the buyers' trust[5].

Analyzing eBay's feedback mechanism: summary of empirical evidence

eBay's impressive commercial success seems to indicate that its feedback mechanism has succeeded in achieving its primary objective: generate sufficient trust among buyers to persuade them to assume the risk of transacting with complete strangers. The sustained growth of eBay's community can only mean that the same mechanism has also succeeded in persuading sellers to behave sufficiently well towards buyers.

Since *sufficiently* does not necessarily mean *efficiently*, eBay's success has generated substantial interest in understanding how well its feedback mechanism works and how its success can be replicated in other environments.

---

[5] eBay transactions are also vulnerable to misbehavior on the part of the *buyer*, most notably situations where a buyer wins an auction but fails to send payment to the seller. It is for that reason that eBay allows sellers to rate buyers as well. To simplify the discussion, in the rest of this section we will assume that buyers are well behaved and will concentrate our attention on seller behavior, with the understanding that the impact of eBay's feedback mechanism on buyer behavior is an area that has so far not received enough attention and requires further study.

A first set of useful results comes from empirical studies. Even a surface analysis of a representative eBay data set can uncover some interesting properties (Resnick and Zeckhauser, 2001):

- Most trading relationships are one-time deals: 89% of all buyer-seller pairs conducted just one transaction during the five-month period covered by the data set
- Buyers left feedback on sellers 52.1% of the time; sellers on buyers 60.6% of the time
- Feedback is overwhelmingly positive; of feedback provided by buyers 99.1% of comments were positive, 0.6% were negative and 0.3% were neutral

A number of studies have delved deeper into eBay data sets in order to uncover additional properties. Resnick, Zeckhauser, Swanson and Lockwood (2002) provide a comprehensive survey and methodological critique of these works. Table 2, based on their survey, summarizes the main findings of these studies. Table 3 provides an alternative summary, focusing on the questions that these studies have addressed. The following points summarize the main conclusions derived from a collective reading of these works:

- Feedback profiles seem to affect both prices and the probability of sale. However, the precise effects are ambivalent; different studies focus on different components of eBay's complex feedback profile and often reach different conclusions. In almost all cases, the quantitative impact of feedback profiles on prices and probability of sale, although statistically significant, is relatively mild.
- The impact of feedback profiles on prices and probability of sale is relatively higher for riskier transactions and more expensive products.
- Among all different pieces of feedback information that eBay publishes for a member (Figures 1 and 2), the components that seem to be most influential in affecting buyer behavior are the overall number of positive and negative ratings, followed by the number of recently (last 7 days, last month) posted negative comments.

Shorthand	Citation	Items sold	Remarks
BP	Ba and Pavlou, 2002	Music, Software, Electronics	Positive feedback increased estimated price, but negative feedback did not have an effect
BH	Bajari and Hortascu, 2000	Coins	Both positive and negative feedback affect probability of modeled buyer entry into the auction, but only positive feedback had a significant effect on final price
DH	Dewan and Hsu, 2001	Stamps	Higher net score increases price
E	Eaton, 2002	Electric guitars	Negative feedback reduces probability of sale, but not price of sold items
HW	Houser and Wooders, 2000	Pentium chips	Positive feedback increases price; negative feedback reduces it
KM	Kalyanam and McIntyre, 2001	Palm Pilot PDAs	Positive feedback increases price; negative feedback reduces price
KW	Kauffman and Wood, 2000	Coins	No significant effects, but negative feedback seems to increase price (!) in univariate analysis
LIL	Lee, Im and Lee, 2000	Computer monitors and printers	Negative feedback reduces price, but only for used items
L	Livingston, 2002	Golf clubs	Positive feedback increases both likelihood of sale and price; effect tapers off once a record is established
LBPD	Lucking-Reiley et. al., 2000	Coins	No effect from positive feedback; negative feedback reduces price
MA	Melnik and Alm, 2002	Gold coins	Positive feedback increases price; negative feedback decreases price
MS	McDonald and Slawson, 2000	Dolls	Higher net score (positives -negatives) increases price
RZ	Resnick and Zeckhauser, 2002	MP3 players, Beanie babies	Both forms of feedback affect probability of sale but not price contingent on sale
RZSL	Resnick Zeckhauser, Swanson and Lockwood, 2002	Vintage postcards	Controlled field experiment; established seller commands higher prices than newcomers; among newcomers, small amounts of negative feedback have little effect

Table 2 Empirical studies on eBay: summary of results [6]

Question considered	Studies
How does a seller's feedback profile affect prices?	all
How does a seller's feedback profile affect the probability of sale?	BH, E, L, RZ
Does feedback matter more for riskier transactions/more expensive products?	BP, LIL
How do prices on eBay compare to prices in a more conventional channel?	DH, KW
What components of eBay's feedback profile better explain buyer behavior?	DH

Table 3 Empirical studies on eBay: summary of questions

Towards a systematic discipline of reputation mechanism design

The evidence provided by this initial group of empirical studies, though useful, does not directly answer the most important underlying question: how well does eBay's mechanism work? In fact,

[6] Adapted and expanded from (Resnick, Zeckhauser, Swanson and Lockwood, 2002).

the findings of these studies raise a whole new set of questions (as they should). For example, why is the fraction of negative feedback so low? Is this an indication of the mechanism's poor functioning (buyers are reluctant to express their true opinions) or perhaps a consequence of the mechanism's success (sellers are induced to behave well and therefore, there simply are very few dissatisfied buyers)? Why is the relationship between feedback and prices so ambivalent? Is this an indication that users do not pay attention to the mechanism, or perhaps it shows that users trust the mechanism so much that they discount small fluctuations in a seller's profile as noise?

In the author's opinion the two most concrete evaluation criteria of a reputation mechanism's performance ought to be (a) the expected payoffs of the outcomes induced by the mechanism for the various classes of stakeholders over the entire time horizon that matters for each of them, and (b) the robustness of those outcomes against different assumptions about the participants' behavior[7].

Calculation of payoffs requires an understanding of how eBay's mechanism affects the bidding behavior of buyers and the pre- and post-auction behavior of sellers and how these behaviors evolve over time. The tools of game theory are instrumental in developing conceptual models of such behavior. Further theory-driven simulation, empirical and experimental studies are, however, also essential, both for qualifying these models, as well as for adapting them to account for the bounded rationality of actual human behavior[8].

Robustness considerations are especially important on eBay since the whole concept of reputation relies on voluntary elicitation of behavior and this, in turn, relies on a number of assumptions about human rationality and beliefs. Two issues stand out as particularly important: First, since feedback provision is currently voluntary, the impact of incomplete and/or untruthful feedback needs to be better understood. Second, the vulnerability of the system against strategic manipulation and online identity changes must be carefully studied.

Once we have sufficiently understood the properties and performance of eBay's current mechanism, the next obvious question is: how can this mechanism be improved? The answer to

[7] Other plausible, but currently less well understood evaluation criteria include inducing outcomes that are perceived as "fair" by the majority of players and ensuring the privacy of participants (Shoham, 2002).
[8] See (Roth, 2002) for a broad discussion of the new methodological challenges introduced by the increasing use of economics not only for analyzing markets but also for designing them.

this question requires a better understanding of the unique design possibilities of online reputation mechanisms. Here are a few examples of a much larger set of possibilities:

- Online reputation mechanisms can precisely control the form of information they solicit: eBay asks users to rate transactions as “positive”, “negative” or “neutral”. Would it have been better to maybe ask them to rate transactions on a scale from 1-5? Would some other ways to phrase the questions lead to even higher efficiency?

- Reputation mechanisms control how information gets aggregated and what information is publicly available in feedback profiles. Currently, eBay’s feedback profile is a relatively complex artifact that includes the entire history of ratings together with a number of summary statistics (Figures 1 and 2). Since different users pay attention to different subsets of this information, this complicates the modeling and predictions of the induced outcomes. Would it be better to hide some parts of this information (for example, the detailed feedback history)? Would inclusion of some other summary statistics (e.g. the *fraction* of negative ratings) lead to even more efficient outcomes? Would it make sense to implement some sort of automated filtering of ratings that fail to satisfy some criteria?

- Feedback submission is currently voluntary on eBay. Furthermore, there is currently no quality control of submitted feedback. Could eBay introduce a carefully designed system of buyer fees and rewards that elicits complete participation and truthful feedback?

Finally, a third set of questions revolve around how online reputation mechanisms compare against more established institutions for achieving similar outcomes, such as formal contracts and brand-building. These comparisons are important; their outcome will help determine how wide of an impact these mechanisms will ultimately have in our society.

An objective of any discipline of design is to eventually be able to abstract from the study of specific cases and articulate some general principles and guidelines. In the case of reputation mechanisms this objective translates to recognizing general classes of settings where reputation mechanisms may be usefully applied, identifying important families of reputation mechanism architectures and understanding what architectures are best suited to what settings.

The rest of the paper provides a survey of past work that can serve as a starting point for answering the above questions in a systematic way. Since the study of reputation mechanisms has not yet been recognized as a field of its own, these results come from a variety of disciplines including economics, information systems, artificial intelligence and psychology.

4 REPUTATION IN GAME THEORY AND ECONOMICS

Given the importance of word-of-mouth networks in human society, reputation formation has been extensively studied by economists using the tools of game theory. This body of work is perhaps the most promising foundation for developing an analytical discipline of online reputation mechanism design. This section surveys past work in this area, emphasizing the results that are most relevant to the design of online reputation mechanisms. Section 5 then discusses how this body of work is being extended to address the unique properties of online systems.

4.1 Basic Concepts

According to Wilson (1985) reputation is a concept that arises in repeated game settings when there is uncertainty about some property (the "type") of one or more players in the mind of other players. If "uninformed" players have access to the history of past stage game outcomes, reputation effects then often allow informed players to improve their long-term payoffs by gradually convincing uninformed players that they belong to the type that best suits their interests. They do this by repeatedly choosing actions that make them appear to uninformed players as if they were of the intended type (thus "acquiring a reputation" for being of that type).

The existence of some initial doubt in the mind of uninformed players regarding the type of informed players is crucial in order for reputation effects to occur. To see this, consider a repeated game between a long-run player and a sequence of short-run (one-shot) opponents. In every stage game the long-run player can choose one out of several actions but cannot credibly commit to any of those actions in advance. If there is no uncertainty about the long-run player's type[9], rational short-run players will then always play their stage-game Nash equilibrium response. Such behavior typically results in inefficient outcomes.

For example, consider an eBay seller who faces an infinite sequence of sets of identical one-time buyers in a marketplace where there are only two kinds of products: high-quality products that cost 0 to the seller and are worth 1 to the buyers and low-quality products that cost 1 to the seller and are worth 3 to the buyers. Buyers compete with one another on a Vickrey auction and therefore bid amounts equal to their expected valuation of the transaction outcome. The winning bidder sends payment to the seller and the seller then has the choice of either "cooperating" (producing a high quality good) or "cheating" (producing a low quality good). The resulting payoff matrix is depicted in Table 4. If the seller cannot credibly pre-commit to cooperation, the expected outcome of all stage games will be the socially inefficient Nash equilibrium: sellers will always cheat and, expecting this, buyers always place low bids.

	Cooperate	Cheat
Bid high	0,2	-2,3
Bid low	2,0	0,1

Table 4 Payoff matrix of a simplified "eBay" bilateral exchange stage game

The concept of reputation allows the long-run player to improve his payoffs in such settings. Intuitively, a long-run player who has a track record of playing a given action (e.g. cooperate) often enough in the past acquires a reputation for doing so and is "trusted" by subsequent short-run players to do so in the future as well. However, why would a profit-maximizing long-term player be willing to behave in such a way and why would rational short-term players use past history as an indication of future behavior?

To explain such phenomena, Kreps, Milgrom, Roberts and Wilson (1982), Kreps and Wilson (1982), and Milgrom and Roberts (1982) introduced the notion of "commitment" types. Commitment types are long-run players who are locked into playing the same action[10]. An

[9] In other words, if short-run players are convinced that the long-run player is a rational utility-maximizing player whose stage-game payoffs are known with certainty.
[10] Commitment types are sometimes also referred to as "irrational" types because they follow fixed, "hard-wired" strategies as opposed to "rational" profit-maximizing strategies. An alternative way to justify such players is to consider them as players with non-standard payoff structures such that that the "commitment" action is their dominant strategy given their payoffs.

important subclass of commitment types are Stackelberg types: long-run players who are locked into playing the, so called, Stackelberg action. The Stackelberg action is the action to which the long-run player would credibly commit if he could. In the above "eBay" example the Stackelberg action would be to cooperate; cooperation is the action that maximizes the seller's lifetime payoffs if the seller could credibly commit to an action for the entire duration of the game. Therefore, the Stackelberg type in this example corresponds to an "honest" seller who never cheats. In contrast, an "ordinary" or "strategic" type corresponds to a profit-maximizing seller who cheats whenever it is advantageous for him to do so.

Reputation models assume that short-run players know that commitment types exist, but are ignorant of the type of the player they face. An additional assumption is that short-run players have access to the entire history of past stage game outcomes[11]. A player's reputation at any given time then consists of the conditional posterior probabilities over that player's type given a short-run player's prior probabilities over types and the repeated application of Bayes' rule on the history of past stage game outcomes.

In such a setting, when selecting his next move the informed player must take into account not only his short-term payoff, but also the long-term consequences of his action based on what that action reveals about his type to other players. As long as the promised future gains due to the increased (or sustained) reputation that comes from playing the Stackelberg action offset whatever short-term incentives he might have to play otherwise, the equilibrium strategy for an "ordinary" informed player will be to try to "acquire a reputation" by masquerading as a Stackelberg type (i.e. repeatedly play the Stackelberg action with high probability.)

In the "eBay" example, if the promised future gains of reputation effects are high enough[12], ordinary sellers are induced to overcome their short-term temptation to cheat and to try to acquire a reputation for honesty by repeatedly producing high quality. Expecting this, buyers will then place high bids, thus increasing the seller's long-term payoffs.

---

[11] The traditional justification for this assumption is that past outcomes are either publicly observable or explicitly communicated among short-run players. The emergence of online feedback mechanisms provides, of course, yet another justification (however, the *private* observability of outcomes in online systems introduces a number of complications; see Section 5.2).

[12] In this type of game this requires that (a) the remaining horizon of the seller is long enough, and (b) the profit margin of high quality products is high enough relative to the discount factor (see Shapiro, 1983).

4.2 Who benefits from reputation?

In general, reputation effects benefit the most patient player in the game: the player who discounts future payoffs less is usually the one who is able to reap the benefits of reputation. This effect is best understood in repeated games where a long-run player faces a sequence of one-shot opponents. This setting is very relevant to online communities since the huge number of participants in such environments makes repeated interaction between the same players unlikely. In such cases, reputation allows the long-run player to masquerade as the commitment type of his choice and "force" short-run players to play a best response to his apparent type. Fudenberg and Levine (1992) show that this result holds even when players can observe only noisy signals of each other's actions, so that the game has imperfect public monitoring. They prove that, if short-run players assign positive prior probability to the long-run player being a Stackelberg type and if that player is sufficiently patient, then an ordinary long-run player achieves an average discounted payoff close to his commitment payoff (i.e., his payoff if he could credibly commit to the Stackelberg action). In order to obtain this payoff, the ordinary player spends long periods of time choosing the Stackelberg action with high probability[13].

Facing longer-lived opponents may be worse for the informed player and generally results in less sharp predictions about reputation effects (Cripps and Thomas, 1995; Cripps, Schmidt and Thomas, 1996). Quite interestingly, however, in repeated games where a patient player faces one or more long-lived but less patient opponents, if the more patient player does not observe the less patient players' intended actions but only sees an imperfect signal of them, reputation effects once again become strong and result in lower bounds that are even higher than in the case where all opponents are myopic (Celentani, Fudenberg, Levine and Pesendorfer, 1996). This last case is equivalent to a situation where a long-run player faces a sequence of long-run but "infrequent" players. This is, perhaps, an even more realistic model of relationships in online communities and therefore an area that deserves further study.

A corollary of the above discussion is that reputation phenomena benefit specific categories of players and are not necessarily socially beneficial. In the above "eBay" example reputation effects happen to be socially beneficial (more precisely, they benefit the seller without harming

[13] This result also requires that the stage game is either a simultaneous move game, or, in a sequential move game, that the short-run players always observe whether or not the Stackelberg strategy has been played.

the buyers). However, in other games reputation effects can increase or reduce social efficiency depending on the structure of the payoffs[14]. This is an important point both for social planners considering the introduction of online reputation mechanisms in a given setting as well as community members considering whether to participate in them.

4.3 Reputation dynamics

The establishment of bounds on long-term player payoffs in a reputation game provides a high-level characterization of reputation effects but leaves a lot of important questions unanswered. It provides vague predictions of payoffs when discount factors are less than one, gives little information about the fate of short-run players and says nothing about how the equilibrium strategies of players evolve over time.

The derivation of equilibrium strategies in repeated games with reputation effects is, in general, quite complicated. Nevertheless, a small number of specific cases have been extensively studied. They provide interesting insight into the complex behavioral dynamics introduced by reputational considerations.

Initial phase

In most cases, reputation effects begin to work immediately and in fact, are strongest during the initial phase, when players must work hard to establish a reputation. Holmstrom (1999) discusses an interesting model of reputational considerations in the context of an agent's "career" concerns: suppose that wages are a function of an employee's innate ability for a task. Employers cannot directly observe an employee's ability, however, they can keep track of the average value of her past task outputs. Outputs depend both on ability and labor. The employee's objective is to maximize her lifetime wages while minimizing the labor she has to put in. At equilibrium, this provides incentives to the employee to work hard right from the beginning of her career in order

[14] A well-known example of a game where reputation effects usually turn out to be socially harmful is the "chain store" game (Kreps and Wilson, 1982; Milgrom and Roberts, 1982). In this game reputation effects allow an incumbent firm to build a reputation for being "tough" and therefore to deter potential entrants from entering its market. Such effects would allow, say, a monopolist to defend its position as the sole firm in a market and, in most cases, would result in reduced social welfare relative to the case where reputation effects are not present (in that case the expected stage-game Nash equilibrium outcome would be that new firms would enter and the incumbent firm would accommodate them).

to build a reputation for competence. In fact these incentives are strongest at the very beginning of her career when observations are most informative.

During the initial phase of a repeated game it is common that some players realize lower, or even negative profits, while the community "learns" their type. In those cases players will only attempt to build a reputation if the losses from masquerading as a Stackelberg type in the current round are offset by the present value of the gains from their improved reputation in the later part of the game. In trading environments, this condition usually translates to the need of sufficiently high profit margins for "good quality" products in order for reputation effects to work. This was first pointed out in (Klein and Leffler, 1981) and explored more formally in (Shapiro, 1983).

Another case where reputation effects may fail to work is when short-run players are "too cautious" vis-à-vis the long-run player and therefore update their beliefs too slowly in order for the long-run player to find it profitable to try to build a reputation. Such cases may occur when, in addition to Stackelberg ("good") types the set of commitment types also includes "bad" or "inept" types: players who always play the action that the short-run players like least. In the "eBay" example, a "bad" type corresponds to a player who always cheats. If short-run players have a substantial prior belief that the long-run player may be a "bad" type then the structure of the game may not allow them to update their beliefs fast enough to make it worthwhile for the long-run player to try to acquire a reputation.

Diamond's (1989) analysis of reputation formation in debt markets presents an example of such a setting. In Diamond's model there are three types of borrowers: safe borrowers, who always select safe projects (i.e. projects with zero probability of default), risky borrowers, who always select risky projects (i.e. projects with higher returns if successful but with nonzero probability of default) and strategic borrowers who will select the type of project that maximizes their long term expected payoff. The objective of lenders is to maximize their long term return by offering competitive interest rates, while at the same time being able to distinguish profitable from unprofitable borrowers. Lenders do not observe a borrower's choice of projects, but they do have access to her history of defaults. In Diamond's model, if lenders believe that the initial fraction of risky borrowers is significant, then, despite the reputation mechanism, at the beginning of the game interest rates will be so high that strategic players have an incentive to select risky projects. Some of them will default and will exit the game. Others will prove lucky and will begin to be considered as safe players. It is only after lucky strategic players have already acquired some initial reputation (and therefore begin to receive lower interest rates) that it becomes optimal for them to begin "masquerading" as safe players by consciously choosing safe projects in order to maintain their good reputation.

Steady state (or lack thereof)

Reputation games are ideally characterized by an equilibrium in which the long-run player repeatedly plays the Stackelberg action with high probability and the player's reputation converges to the Stackelberg type.

The existence of such steady states crucially depends on the ability to perfectly monitor the outcomes of individual stage games. In games with perfect public monitoring of stage game outcomes such a steady state almost always exists. For example, consider the "eBay game" that serves as an example throughout this section, with the added assumption that buyers perfectly and truthfully observe and report the seller's action. In such cases, the presence of even a single negative rating on a seller's feedback history reveals the fact that the seller is not honest. From then on, buyers will always choose the low bid in perpetuity. Since such an outcome is not advantageous for the seller, reputation considerations will induce the seller to cooperate forever.

The situation changes radically if monitoring of outcomes is imperfect. In the eBay example, imperfect monitoring means that even when the seller produces high quality there is a possibility that an eBay buyer will post a negative rating, and, conversely, even when the seller produces low quality, the buyer may post a positive rating. A striking result is that in such "noisy" environments reputations cannot be sustained indefinitely: if a strategic player stays in the game long enough, short-run players will eventually learn his true type and the game will inevitably revert to one of the static Nash equilibria (Cripps, Mailath and Samuelson, 2002).

To see the intuition behind this result, note that reputations under perfect monitoring are typically supported by a trigger strategy. Deviations from the equilibrium strategy reveal the type of the deviator and are punished by a switch to an undesirable equilibrium of the resulting complete-information continuation game. In contrast, when monitoring is imperfect, individual deviations neither completely reveal the deviator's type nor trigger punishments. Instead, the long-run convergence of beliefs ensures that eventually any current signal of play has an arbitrarily small effect on the uniformed player's beliefs. As a result, a player trying to maintain a reputation ultimately incurs virtually no cost (in terms of altered beliefs) from indulging in a single small deviation from Stackelberg play. But the long-run effect of many such small deviations from the commitment strategy is to drive the equilibrium to full revelation.

Holmstrom's "career concerns" paper provides an early special case of this striking result: the longer an employee has been on the market, the more "solid" the track record she has acquired and the less important her current actions in influencing the market's future assessment of her ability. This provides diminishing incentives for her to keep working hard. Cripps, Mailath and Samuelson's result then states that, if the employee stays on the market for a really long time, these dynamics will lead to an eventual loss of her reputation[15].

If one tries to reconcile the above result with Fudenberg and Levine's 1992 result on long-term payoffs induced by reputation, even more interesting phenomena come to the surface: If players eventually lose their reputation, in order for them to achieve average long-term payoffs that are close to their Stackelberg payoff, they must realize payoffs higher than their Stackelberg payoff during some stages of the game. This makes the dynamics of reputation formation in environments with imperfect monitoring quite complex indeed: an initial phase of reputation formation (with potentially suboptimal payoffs) is followed by a phase where the long-run player has established a reputation and is able to occasionally "fool" short-run players (thus realizing payoffs above his Stackelberg payoff), followed by a phase where short-run players eventually learn the truth and the game reverts to its static stage game Nash equilibrium.

These dynamics have important repercussions for systems like eBay. Dellarocas (2002c) has conducted simulation studies of the dynamics induced by eBay's mechanism in settings with imperfect monitoring and moral hazard. He finds that if eBay makes the entire feedback history of a seller (component D in Figure 2) available to buyers and if an eBay seller stays on the system long enough, once he establishes an initial reputation for honesty he will be tempted to occasionally cheat buyers. In the long run, this behavior will lead to an eventual collapse of his reputation and therefore of cooperative behavior. The conclusion is that, if buyers pay attention

[15] Another interesting study of a situation exhibiting similar dynamics is Benabou and Laroque's 1992 model of market insiders who have incentives to manipulate public information and asset prices through strategically distorted announcements or forecasts.

to a seller's entire feedback history, eBay's current mechanism fails to sustain long-term cooperation.

Endgame considerations

Since reputation relies on a tradeoff between current "restraint" and the promise of future gains, in finitely repeated games incentives to maintain a reputation diminish and eventually disappear as the end of the game comes close.

One solution to this problem is to introduce community membership rules that elicit good behavior throughout the game (Ba, 2001). For example, online communities can levy a sufficiently high entrance fee that is refundable subject to maintaining a good reputation upon exit.

Another solution is to assign some post-mortem value to reputation, so that players find it optimal to maintain it throughout the game. For example, reputations can be viewed as assets that can be bought and sold in a market for reputations. Tadelis (1998) shows that a market for reputations is indeed sustainable. Furthermore, the existence of such a market provides "old" agents and "young" agents with equal incentives to exert effort (Tadelis, 2002). However, the long-run effects of introducing such a market can be quite complicated since good reputations are then likely to be purchased by "inept" agents for the purpose of depleting them (Mailath and Samuelson, 2001; Tadelis, 2002). Further research is needed in order to fully understand the long-term consequences of introducing markets for reputations as well as for transferring these promising concepts to the online domain.

4.4 Robustness of reputation models

The fundamental assumption underlying most game theoretic models of reputation is that all players have identical prior beliefs and that behavior is consistent with the concept of Bayesian Nash equilibrium. These assumptions are probably too stringent and unrealistic in environments as diverse as large-scale online communities. Fortunately, reputation phenomena arise under significantly weaker assumptions on the knowledge and behavior of players. Watson (1993; 1996) and Battigalli and Watson (1997) demonstrated that reputation effects do not require equilibrium. They are implied by a weak notion of rationalizability along with two main conditions on the beliefs of players: First, there must be a strictly positive and uniform lower bound on the subjective probability that players assign to the Stackelberg type. Second, the conditional beliefs of short run players must not be "too dispersed".

Even though the *emergence* of reputation effects is quite robust, their dynamics and resulting payoffs tend to be quite sensitive to the distribution of prior beliefs on the existence of commitment types. In Diamond's model, for example, the introduction of "inept" types alongside Stackelberg types leads to failure of reputation effects during the initial phase. Mailath and Samuelson (1998) challenge the justifiability of Stackelberg types in many real-life settings. They show that if this assumption is replaced with the assumption of an inept type, substantially different effects emerge. Ely and Valimaki (2002) construct a striking example where the assumption of specific "bad" commitment types leads to situation where reputation effects end up *reducing* the long-run players' payoffs[16].

In the opinion of the author this is a caveat of reputation models that has important implications for designers of online reputation mechanisms: Reputation mechanisms should ideally be designed to be as robust as possible against different prior beliefs on commitment types. If this is not possible then online community operators can try to structure the "look-and-feel" and other "intangible" aspects of their systems so as to "steer" the predispositions of community members in the direction that makes reputation mechanisms most effective. This argument provides an economic justification for eBay's obvious efforts to instill a culture of optimism, cheerfulness and praise among its members (Resnick and Zeckhauser, 2001). Theory predicts that positive predispositions (prior beliefs) of buyers towards sellers help make reputation effects stronger, therefore inducing sellers to behave better, and finally resulting in higher long-term gains for all parties involved.

5 New opportunities and challenges of online Mechanisms

In Section 2, I discussed a number of differences between online reputation mechanisms and traditional word-of-mouth networks. This section surveys our progress in understanding the opportunities and challenges that these special properties imply.

[16] Their result was generalized by Ely, Fudenberg and Levine (2002).

5.1 Understanding the impact of scalability

The impact of the vastly increased scale of online reputation mechanisms relative to their brick-and-mortar counterparts is not yet fully understood. Bakos and Dellarocas (2002) model the impact of information technology on online reputation mechanisms in the context of a comparison of the social efficiency of litigation and online reputation. They observe that online reputation mechanisms provide linkages between otherwise disconnected smaller markets (each having its own informal word-of-mouth networks) in which a firm operates. This, in turn, is equivalent to increasing the discount factor of the firm when it considers the future impacts of its behavior on any given transaction. In trading relationships, a minimum discount factor is necessary to make reputation effects effective *at all* in inducing cooperative behavior (this is an alternative way to interpret Klein and Leffler, 1981 and Shapiro, 1983). Above that threshold, higher discount factors result in higher efficiency. Bakos and Dellarocas show how, under certain conditions, sufficiently large reputation mechanisms can be a more socially efficient institution for inducing honest trade than the threat of litigation.

5.2 Eliciting sufficient and honest feedback

Most game theoretic models of reputation formation assume that stage game outcomes (or imperfect signals thereof) are *publicly* observed. Online reputation mechanisms, in contrast, rely on *private* monitoring of stage game outcomes and voluntary feedback submission. This introduces two important new considerations (a) ensuring that sufficient feedback is, indeed, provided and (b) inducing truthful reporting.

Economic theory predicts that voluntary feedback will be underprovided. There are two main reasons for this. First, feedback constitutes a public good: once available, everyone can costlessly benefit from it. Voluntary provision of feedback leads to suboptimal supply, since no individual takes account of the benefits that her provision gives to others. Second, provision of feedback presupposes that the rater will assume the risks of transacting. Such risks are highest for new products: prospective consumers may be tempted to wait until more information is available. However, unless somebody decides to take the risk of becoming an early evaluator, no feedback will ever be provided.

Avery, Resnick and Zeckhauser (1999) analyze mechanisms whereby early evaluators are paid to provide information and later evaluators pay so as to balance the budget. They conclude that any two of three desirable properties for such a mechanism can be achieved, but not all three, the three properties being voluntary participation, no price discrimination and budget balance.

Since monitoring is private and assessments usually subjective, an additional consideration is whether feedback is honest. Miller, Resnick and Zeckhauser (2002) propose a mechanism for eliciting honest feedback based on the technique of proper scoring rules. A scoring rule is a method for inducing decision makers to reveal their true beliefs about the distribution of a random variable by rewarding them based on the actual realization of the random variable and their announced distribution (Cooke, 1991). A proper scoring rule has the property that the decision maker maximizes the expected score when he truthfully announces his belief about the distribution.

Their mechanism works as long as raters are assumed to act independently. Collusive behavior can defeat proper scoring rules. Unfortunately, online environments are particularly vulnerable to collusion. The development of effective mechanisms for dealing with collusive efforts to manipulate online ratings is currently an active area of research. Dellarocas (2000; 2001) explores the use of robust statistics in aggregating individual ratings as a mechanism for reducing the effects of coordinated efforts to bias ratings. To this date, however, there is no effective solution that completely eliminates the problem.

5.3 Exploiting the information processing capabilities of feedback mediators

Most game theoretic models of reputation assume that short-run players have access to the entire past history of stage game outcomes and update their prior beliefs by repeated application of Bayes' rule on that information.

Online feedback mediators completely control the amount and type of information that is made available to short-run players. This opens an entire range of new possibilities: For example, feedback mediators can hide the detailed history of past feedback from short-term players and replace it with a summary statistic (such as the sum, mean or median of past ratings) or with any other function of the feedback history. They can filter outlying or otherwise suspect ratings. They can offer *personalized* feedback profiles, that is, present different information about the same long-run player to different short-run players.

Such information transformations can have non-trivial effects in the resulting equilibria and can allow online reputation mechanisms to induce outcomes that are difficult or impossible to attain in standard settings. The following are two examples of what can be achieved:

As discussed in Section 4.3, in environments with imperfect monitoring traditional reputation models predict that reputations are not sustainable: once firms build a reputation they are tempted to "rest on the laurels"; this behavior, ultimately, leads to a loss of their reputation. Economists have used a variety of devices to construct models that do not exhibit this undesirable behavior. For instance, Mailath and Samuelson (1998) assume that in every period there is a fixed, exogenous probability that the type of the firm might change. Horner (2002) proposes a model in which competition among firms induces them to exert sustained effort.

Online feedback mediators provide yet another, perhaps much more tangible approach to eliminating such problems: by *designing* the mediator to only publish recent feedback, firms are given incentives to constantly exert high effort. In the context of eBay, this result argues for the elimination of the detailed feedback history and the use of summary statistics as the primary focal point of feedback profiles. Dellarocas (2002a) studies the equilibria induced by a variation of eBay's feedback mechanism in which the only information available to short-run players is the sum of positive and negative ratings posted on a seller during the most recent N transactions. He finds that, unlike the case where a seller's entire history is published, such a mechanism induces high levels of cooperation that do not decline over time. Furthermore, the maximum long-run payoffs are independent of the size of the window N: a mechanism that only publishes the *single* most recent rating is capable of inducing the same maximum efficiency as one that publishes summaries of arbitrarily large numbers of past ratings.

The difference between the above "information engineering" approach and the typical fashion in which economists attempt to justify the emergence of similar outcomes is that the latter usually rely on justifying *different interpretations of the available information* whereas the former simply uses the power of information technology in order to *change the nature of available information*.

A second example of improving efficiency through proper feedback mediator design can be found in (Dellarocas, 2002b). Dellarocas studies settings in which a monopolist seller sells products of various qualities and announces the quality of each product. The objective of a reputation mechanism in such settings is to induce truthful announcements. Once again, Cripps, Mailath and Samuelson's result predicts that, in noisy environments, a mechanism that simply publishes the entire history of feedback will not lead to sustainable truth telling. Dellarocas proposes a mechanism that acts as an intermediary between the seller and the buyers. The mechanism does not publish the history of past ratings. Instead, it internally keeps track of discrepancies between past seller quality announcements and corresponding buyer feedback. It then punishes or rewards the seller by "distorting" the seller's subsequent quality announcements so as to charge/compensate him for whatever "unfair" gains or losses he has realized by misrepresenting the quality of his items in past rounds. If consumers are risk-averse, at equilibrium this mechanism induces the seller to truthfully announce quality throughout the (infinite version of the) game.

The above examples have only scratched the surface of what is possible to achieve through careful feedback mediator design. Understanding the full range of possibilities introduced by the ability to design the type of information that flows in and out of these systems is one of the most interesting questions in the field of online reputation mechanism design.

5.4 Coping with easy name changes

In online communities it is usually easy for members to disappear and re-register under a completely different online identity with zero or very low cost. Friedman and Resnick (2001) refer to this property as "cheap pseudonyms". This property hinders the effectiveness of reputation mechanisms: community members can build a reputation, milk it by cheating other members and then vanish and re-enter the community with a new identity and a clean record.

Friedman and Resnick discuss two classes of approaches to this issue: Either make it more difficult to change online identities, or structure the community in such a way so that exit and re-entry with a new identity becomes unprofitable. The first approach makes use of cryptographic authentication technologies and is outside the scope of this paper. The second approach is based on imposing an upfront cost to each new entrant, such that the benefits of "milking" one's reputation are exceeded by the cost of subsequent re-entry. This cost can be an explicit entrance fee or an implicit cost of having to go through an initial reputation-building (or *"dues paying"*) phase with low or negative profits. Friedman and Resnick show that, although dues paying approaches incur efficiency losses, such losses constitute an inevitable consequence of easy name changes.

Dellarocas (2002a) shows how such a "dues paying" approach can be implemented in an eBay-like environment where feedback mediators only publish the sum of recent ratings. He proves that, in the presence of easy name changes, the design that results in optimal social efficiency is one where the mechanism sets the initial profile of new members to correspond to the "worst" possible reputation[17]. He further shows that, although this design incurs efficiency losses relative to the case where identity changes are not possible, its efficiency is the highest possible attainable by any mechanism if players can costlessly change their identities.

5.5 Assisting the interpretation of subjective information

Feedback information is strongly influenced by subjective factors such as the rater's tastes and cultural background. In brick-and-mortar interactions recipients of such information rely on a variety of social cues (such as previous experience with the rater or inferences based on the rater's appearance, age, race, social status, etc.) in order to interpret it and "translate" it to their own (subjective) value system. Such cues are usually absent from online communities.

Technology can assist the interpretation of subjective information through the concept of *collaborative filtering* (Resnick et. al., 1994; Shardanand and Maes, 1995; Bresee et. al., 1998). Collaborative filtering techniques rely on the assumption that human communities consist of a relatively small set of "taste clusters": groups of people with similar viewpoints for similar things. If these taste clusters can be identified, then a reputation mechanism can usefully *personalize* the feedback information it distributes to a community member *A* about some other member *B* by only including feedback submitted regarding *B* by members that belong to the same cluster as *A* (or, more generally, by weighting feedback in inverse proportion to the "taste

[17] For example, if the mechanism summarizes the most recent 10 ratings, newcomers would begin the game with a profile that indicates that all 10 recent ratings were negative. An additional assumption is that buyers cannot tell how long a given seller has been on the market and therefore cannot distinguish between newcomers with "artificially tarnished" profiles and dishonest players who have genuinely accumulated many negative ratings.

cluster distance" between *A* and the authors of feedback). Such personalized feedback summaries are still subjective but, in theory at least, easier to interpret because they consist of opinions of people who "think like" their intended reader.

The problem of identifying the "right" taste cluster for a given community member reduces to the well-studied problem of classification/data clustering (Jain, Murty and Flynn, 1999). The similarity of two members is usually a function of the distance of their past ratings for identical items.

Collaborative filtering techniques have received wide attention and are used in a number of commercial reputation mechanisms, such as Amazon.com. Nevertheless, their operation is based on heuristics and, until today, it has been difficult to quantify their impact on social outcomes.

5.6 Exploring alternative architectures

The preceding discussion has assumed a centralized architecture in which feedback is explicitly provided and a single trusted mediator controls feedback aggregation and distribution. Though the design possibilities of even that simple architecture are not yet fully understood, centralized reputation mechanisms do not nearly exhaust the new possibilities offered by information technology.

In recent years the field of multi-agent systems (Jennings, Sycara and Wooldridge, 1998) has been actively researching online reputation systems as a technology for building trust and inducing good behavior in artificial societies of software agents. Two lines of investigation stand out as particularly novel and promising:

Reputation formation based on analysis of "implicit feedback". In our networked society, several traces of an agent's activities can be found on publicly accessible databases. Instead of (or in addition to) relying on explicitly provided feedback, automated reputation mechanisms can then potentially infer aspects of an agent's attributes, social standing and past behavior through collection and analysis of such "implicit feedback" information.

Perhaps the most successful application of this approach to date is exemplified by the Google search engine. Google assigns a measure of reputation to each web page that matches the keywords of a search request. It then uses that measure in order to rank order search hits. Google's page reputation measure is based on the number of links that point to a page, the number of links that point to the pointing page, and so on (Brin and Page, 1998). The underlying assumption is that if enough people consider a page to be important enough in order to place links to that page from their pages, and if the pointing pages are "reputable" themselves, then the information contained on the target page is likely to be valuable. Google's success in returning relevant results is testimony to the promise of that approach.

Pujol , Sangüesa and Delgado (2002) propose a generalization of the above algorithm that "extracts" the reputation of nodes in a general class of social networks. Sabater and Sierra (2002) describe how direct experience, explicit and implicit feedback can be combined into a single reputation mechanism.

Basing reputation formation on implicit information is a promising solution to problems of eliciting sufficient and truthful feedback. Careful modeling of the benefits and limitations of this approach is needed in order to determine in what settings it might be a viable substitute or complement of voluntary feedback provision.

Decentralized reputation architectures. Our discussion of reputation mechanisms has so far implicitly assumed the honesty of feedback mediators. Alas, mediators are also designed and operated by parties whose interests may sometimes diverge from those of community participants.

Decentralizing the sources of reputation is a promising approach for achieving robustness in the presence of potentially dishonest mediators and privacy concerns. A number of decentralized reputation mechanisms have recently been proposed (Zacharia, Moukas and Maes, 2000; Mui, Szolovits and Ang, 2001; Sen and Sajja, 2002; Yu and Singh, 2002). Though novel and intriguing, none of these works provides a rigorous analysis of the behavior induced by the proposed mechanisms or an explicit discussion of their advantages relative to other alternatives. More collaboration is needed in this promising direction between computer scientists, who better understand the new possibilities offered by technology, and social scientists, who better understand the tools for evaluating the potential impact of these new systems.

5.7 Accounting for bounded rationality

The ambition of a discipline of online reputation mechanism design is to be able to engineer social outcomes with a degree of precision that approaches that of engineering design. This, in turn, requires precise modeling not only of the technological components of those systems but also of the human users.

It is well known by now that human behavior does not exactly conform to the traditional economics assumptions of rational maximization of well-defined utility functions[18]. There are at least three areas where more accurate modeling of human judgment and decision-making are essential to predicting the outcomes induced by a reputation mechanism:

- Modeling how short-term players (e.g. buyers) take into account feedback profiles; understanding how the format of such profiles affects their decision-making
- Modeling how long-term players (e.g. sellers) take into account the existence of the reputation mechanism
- Modeling how short-term players rate following a transaction; understanding how the format and wording of feedback questions affects their responses[19]

A couple of recent laboratory experiments provide some initial insight into human behavior vis-à-vis reputation mechanisms. Bolton, Katok and Ockenfels (2002) compare trading in a market with (automatically generated) feedback to a market without, as well as to a market in which the same people interact with each other repeatedly (partners market). They find that, while the reputation mechanism induces a substantial improvement in trading efficiency, surprisingly, it falls short of the efficiency achieved in the partners market. Kaiser (2002) reports the results of a repeated trust game among strangers with and without the ability to provide feedback. She finds that the presence of a feedback mechanism significantly increases the levels of trust and trustworthiness. Furthermore, efficiency is slightly higher if trading partners are informed of the

[18] For an excellent survey of psychological findings relevant to economics the reader is strongly encouraged to consult (Rabin, 1996).
[19] Research in this area might also benefit from insights from the well-established field of survey research. See, for example, (Tourangeau, Rips and Rasinski, 2000).

entire distribution of each other's previous ratings than if they are informed of each other's most recent rating only.

Further progress in understanding how humans react to reputation mechanisms requires close synergy between analytical modeling, prototyping and empirical and experimental research. The ability to precisely control the technological details of these mechanisms opens new possibilities for targeted empirical/experimental work that proves or disproves specific modeling hypotheses, explores how mediator design might be able to compensate for some of the departures from rationality and leads us to a new level of understanding of how different technological design choices affect the outcomes induced by such mechanisms.

6 CONCLUSIONS

Online reputation mechanisms harness the remarkable ability of the Internet to, not only disseminate, but also collect and aggregate information from large communities at very low cost, in order to artificially construct large-scale word-of-mouth networks. Such networks have historically proven to be effective social control mechanisms in settings where information asymmetries can adversely impact the functioning of a community and where formal contracting is unavailable, unenforceable or prohibitively expensive. They are fast emerging as a promising alternative to more established trust building mechanisms in the digital economy.

The design of such mechanisms can greatly benefit from the insights produced by more than twenty years of economics and game theory research on the topic of reputation. These results need to be extended to take into account the unique new properties of online mechanisms such as their unprecedented scalability, the ability to precisely design the type of feedback information that is solicited and distributed, the volatility of online identities and the relative lack of contextual cues to assist interpretation of what is, essentially, subjective information.

The most important conclusion drawn from this survey is that reputation is a powerful but subtle and complicated concept. Its power to induce cooperation without the need for costly and inefficient enforcement institutions is the basis of its appeal. On the other hand, its effectiveness is often ephemeral and depends on a number of additional tangible and intangible environmental parameters.

In order to translate these initial results into concrete guidance for implementing and participating in effective reputation mechanisms further advances are needed in a number of important areas. The following list contains what the author considers to be the most important open areas of research in reputation mechanism design:

- Scope and explore the design space and limitations of mediated reputation mechanisms. Understand what set of design parameters work best in what settings. Develop formal models of those systems in both monopolistic and competitive settings

- Develop effective solutions to the problems of sufficient participation, easy identity changes and strategic manipulation of online feedback

- Conduct theory-driven experimental and empirical research that sheds more light into buyer and seller behavior vis-à-vis such mechanisms

- Compare the relative efficiency of reputation mechanisms to that of more established mechanisms for dealing with information asymmetries (such as state-backed contractual guarantees and brand-name building) and develop theory-driven guidelines for deciding which set of mechanisms to use when

- Identify new domains where reputation mechanisms can be usefully applied

Online reputation mechanisms attempt to artificially engineer heretofore naturally emerging social phenomena. Through the use of information technology, what had traditionally fallen within the realm of the social sciences is, to a large extent, being transformed into an engineering design problem. The potential to engineer social outcomes through the introduction of carefully crafted information systems is opening a new chapter on the frontiers of information technology. It introduces new methodological challenges that require collaboration between several traditionally distinct disciplines, such as economics, computer science, management science, sociology and psychology, in order to be properly addressed. Our networked societies will benefit from further research in this exciting area.

REFERENCES

Akerlof, G. (1970) The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, pp. 488-500.

Avery, C., Resnick, P. and Zeckhauser, R. (1999) The Market for Evaluations. *American Economics Review*, 89 (3), pp. 564-584.

Ba, S. (2001) Establishing online trust through a community responsibility system. *Decision Support Systems*, 31, pp. 323-336.

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

Bajari, P. and Hortacsu, A. (2000). Winner's Curse, Reserve prices and Endogenous Entry: Empirical Insights from eBay Auctions. Working paper, Department of Economics, Stanford University.

Bakos, Y. and Dellarocas, C. (2002) Cooperation without Enforcement? A Comparative Analysis of Litigation and Online Reputation as Quality Assurance Mechanisms. *Proceedings of the 23rd International Conference on Information Systems*, Barcelona, Spain.

Battigalli, P., and Watson, J. (1997). On "Reputation" Refinements with Heterogeneous Beliefs, *Econometrica* 65, pp. 369–374.

Benabou, R., and G. Laroque (1992) Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility, *Quarterly Journal of Economics*, 107, pp. 921–958.

Benson, B. (1989). The Spontaneous Evolution of Commercial Law. *Southern Economic Journal* 55, pp. 644-61. Reprinted in *Reputation: Studies in the Voluntary Elicitation of Good Conduct*, edited by Daniel B. Klein, pp. 165-89. Ann Arbor: University of Michigan Press, 1997.

Bolton, G. E., Katok, E. and Ockenfels, A. (2002) How Effective are Online Reputation Mechanisms? An Experimental Investigation. Working Paper.

Bresee, J.S., Heckerman, D. and Kadie, C. (1998) Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, pp. 43-52, San Francisco, July 24-26, 1998.

Brin S. and Page L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30 (1-7), pp. 107-117.

Celentani, M., Fudenberg, D., Levine, D., and Pesendorfer, W. (1996). Maintaining a Reputation against a Long Lived Opponent, *Econometrica* 64, pp. 691–704.

Cooke, R. M. (1991) *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, New York, NY.

Cripps, M., Mailath, G. and Samuelson., L. (2002) Imperfect Monitoring and Impermanent Reputations, Working Paper.

Cripps, M., Schmidt, K., and Thomas, J. (1996). Reputation in Perturbed Repeated Games, *Journal of Economic Theory* 69, pp. 387–410.

Cripps, M., and Thomas, J. (1995). Reputation and Commitment in Two-Person Repeated Games without Discounting, *Econometrica* 63, pp. 1401–1419.

Dellarocas, C. (2000) Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. *Proceedings of the 2nd ACM Conference on Electronic Commerce*, Minneapolis, MN, October 17-20, 2000.

Dellarocas. C. (2001) Building Trust On-Line: The Design of Reliable Reputation Mechanisms for Online Trading Communities. Working Paper. MIT Sloan School of Management.

Dellarocas (2002a) Efficiency and Robustness of eBay-like Online Feedback Mechanisms in Environments with Moral Hazard. Working Paper.

Dellarocas (2002b) Goodwill Hunting: An economically efficient online feedback mechanism in environments with variable product quality. *Fourth Workshop on Agent-Mediated Electronic Commerce (AMEC IV),* Bologna, Italy.

Dellarocas (2002c) Understanding the dynamics of eBay's reputation mechanism. Working Paper.

Dewan, S. and Hsu, V. (2001). Trust in Electronic Markets: Price Discovery in Generalist Versus Specialty Online Auctions. Working paper.

Diamond, D. (1989). Reputation Acquisition in Debt Markets. *Journal of Political Economy* 97, 4, pp. 828-862.

Eaton, D. H. (2002). Valuing Information: Evidence from Guitar Auctions on eBay. Working paper, Murray State University.

Ely, J. C., Fudenberg, D. and Levine, D. K. (2002) When is Reputation Bad? Working Paper.

Ely, J. C. and Valimaki, J. (2002) Bad Reputation. Working Paper.

Friedman, E. and Resnick, P. (2001). The Social Cost of Cheap Pseudonyms. *Journal of Economics and Management Strategy* 10 (1), pp. 173-199.

Fudenberg, D., and Levine, D. (1992). Maintaining a Reputation when Strategies are Imperfectly Observed, *Review of Economic Studies* 59, pp. 561–579.

Greif, A. (1993). Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *American Economic Review* 83, pp. 525-548.

Holmstrom, B. (1999). Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies*, 66 (1), pp. 169-182.

Horner, J. (2002) Reputation and Competition, *American Economic Review* 92, pp. 644-663.

Houser, D. and Wooders, J. (2000) Reputation in Auctions: Theory and Evidence from eBay. Working Paper, University of Arizona.

Jain, A. K., Murty, M. N. and Flynn, P.J. (1999) Data clustering: a review. *ACM Computing Surveys*, 31 (3), pp. 264-323.

Jennings, N., Sycara, K., and Wooldridge, M. (1998) A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*, 1 (1), pp. 7-38.

Kaiser, C. Trust and Reputation Building in E-Commerce. Working Paper. IBM T.J. Watson Research Center, Yorktown Heights, NY.

Kalyanam, K. and S. McIntyre (2001). Returns to Reputation in Online Auction markets. Working paper W-RW01-02, Santa Clara, CA, Santa Clara University.

Kauffman, R. J., and C. Wood (2000). Running up the Bid: Modeling Seller Opportunism in Internet Auctions, working paper 00-18, University of Minnesota.

Klein, B. and Leffler, K. (1981) The Role of Market Forces in Assuring Contractual Performance. *Journal of Political Economy*. 89 (4), pp. 615-641.

Klein, D. ed. (1997) *Reputation: Studies in the Voluntary Elicitation of Good Conduct.* University of Michigan Press, Ann Arbor, MI.

Kollock, P. (1999) The Production of Trust in Online Markets. In *Advances in Group Processes* (Vol. 16), eds. E.J. Lawler, M. Macy, S. Thyne, and H.A. Walker, Greenwich, CT: JAI Press.

Kreps, D., Milgrom, P., Roberts, J and Wilson, R. (1982). Rational Cooperation in the Finitely Repeated Prisoners' Dilemma. *Journal of Economic Theory*, 27, pp. 245-52.

Kreps, D. and Wilson, R. (1982). Reputation and Imperfect Information. *Journal of Economic Theory*, 27, pp. 253-79.

Lee, Z., I. Im, et al. (2000). The Effect of Negative Buyer Feedback on Prices in Internet Auction Markets. *Proceedings of the 21st International Conference on Information Systems*, Brisbane, Australia.

Livingston, J. (2002) Can a Contract Be Self-Enforced Using Reputation? A Sample Selection Model of Internet Auctions. Working Paper, University of Maryland.

Lucking-Reiley, D., D. Bryan, et al. (2000). "Pennies from eBay: the Determinants of Price in Online Auctions", Working Paper, Vanderbilt University.

Mailath, G. J., and Samuelson, L. (1998) Your Reputation Is Who You're Not, Not Who You'd Like to Be, Center for Analytic Research in Economics and the Social Sciences (CARESS) Working Paper 98-11, University of Pennsylvania.

Mailath, G. J., and Samuelson, L. (2001) Who Wants a Good Reputation? *Review of Economic Studies,* 68, pp. 415-441.

Mayzlin, D. (2002). Promotional Chat on the Internet. Working Paper, Yale University, 2002.

McDonald, C. G., and V. C. Slawson Jr. (2001). Reputation in an Internet Auction Market. Working paper

Melnik, M. I. and J. Alm (2002). Does a Seller's Reputation Matter? Evidence from eBay Auctions. *Journal of Industrial Economics*, September 2002 (forthcoming).

Milgrom, P. R., North, D., and B. R. Weingast. (1990). The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs. *Economics and Politics* 2, pp. 1-23.

Milgrom, P., and Roberts, J. (1982). Predation, Reputation and Entry Deterrence, *Journal of Economic Theory* 27, pp. 280–312.

Miller, N., Resnick, P., and Zeckhauser, R. (2002) Eliciting Honest Feedback in Electronic Markets, Working Paper.

Mui, L., Szolovits, P. and Ang, C. (2001) Collaborative sanctioning: applications in restaurant recommendations based on reputation. *Proceedings of the fifth international conference on Autonomous agents*, Montreal, Quebec, Canada, pp. 118-119.

Pujol, J. M., Sanguesa, R. and Delgado, J. (2002) Extracting reputation in multi agent systems by means of social network topology. *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, Bologna, Italy, pp. 467-474.

Rabin, M. (1998) Psychology and Economics. *Journal of Economic Literature*, 36, pp. 11-46.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994) Grouplens: An Open Architecture for Collaborative Filtering of Netnews. *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work*, pp. 175-186.

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

Resnick, P., and Zeckhauser, R. (2002) Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In *The Economics of the Internet and E-Commerce*. Michael R. Baye, editor. Volume 11 of *Advances in Applied Microeconomics*. Amsterdam, Elsevier Science (forthcoming).

Resnick, P., Zeckhauser, R., Swanson, J. and Lockwood, K. (2002) The Value of Reputation on eBay: A Controlled Experiment. Working Paper.

Roth, A. (2002) The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. *Econometrica,* 70 (4), pp. 1341-1378

Sabater, J. and Sierra, C. (2002) Reputation and social network analysis in multi-agent systems. *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, Bologna, Italy, pp. 475-482.

Sen, S. and Sajja, N. (2002) Robustness of reputation-based trust: boolean case. *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, Bologna, Italy, pp. 288-293.

Shapiro, C. (1983). Premiums for High Quality Products as Returns to Reputations. *The Quarterly Journal of Economics*, 98 (4), pp. 659-680.

Shardanand, U. and Maes, P. (1995). Social information filtering: Algorithms for automating “word of mouth”. *Proceedings of the Conference on Human Factors in Computing Systems* (CHI95), Denver, CO, pp. 210-217.

Shoham, Y. (2002) It’s not your father’s mechanism design. Oral Presentation.

Tadelis, S. (1999). What's in a Name? Reputation as a Tradeable Asset. *The American Economic Review,* 89 (3), pp. 548-563.

Tadelis, S. (2002). The Market for Reputations as an Incentive Mechanism. *Journal of Political Economy*, 92 (2), pp. 854-882

Tourangeau, R., Rips, L. J. and Rasinski, K. (2000) *The Psychology of Survey Response.* Cambridge University Press, Cambridge, UK.

Watson, J. (1993). A ‘Reputation’ Refinement without Equilibrium, *Econometrica* 61, pp. 199-205.

Watson, J. (1996). Reputation in Repeated Games with No Discounting, *Games and Economic Behavior* 15, pp. 82–109.

Wilson, R. (1985). Reputations in Games and Markets. In *Game-Theoretic Models of Bargaining*, edited by Alvin Roth, Cambridge University Press, pp. 27-62.

Yu, B. and Singh, M. (2002) An evidential model of distributed reputation management. *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, Bologna, Italy, pp. 294-301.

Zacharia, G., Moukas, A. and Maes, P. (2000) Collaborative Reputation Mechanisms in Electronic Marketplaces. *Decision Support Systems*, 29 (4), pp. 371-388.