Co-evolving Trust Mechanisms for Catering User Behavior

Tanja Ažderska

HAL Id: hal-01517663

https://inria.hal.science/hal-01517663

Submitted on 3 May 2017

Co-evolving trust mechanisms for catering user behavior

Tanja Ažderska

Laboratory for Open Systems and Networks, Jozef Stefan Institute, Jamova 39,
1000 Ljubljana, Slovenia

atanja@e5.ijs.si

Abstract. While most of the computational trust models devote to truthfully detecting trustworthy individuals, much less attention is paid to how these models are perceived by users, who are the core of the trust machinery. Understanding the relation between trust models and users' perception of those models may contribute for reducing their complexity, while improving the user-experience and the system performance. Our work recognizes reputation, recommendation and rating systems as online trust representatives and explores the biased behavior resulting from users' perception of those systems. Moreover, we investigate the relation and inter-dependencies between trust mechanisms and user behavior with respect to context, risk, dynamics and privacy. We perform experimental study and identify few types of cognitive biases that users exhibit. Based on the identified factors and the findings of the study, we propose a framework for addressing some of the issues attributed to users' biased behavior.

Keywords: trust, bias, context, reputation, recommendation, rating

1 Introduction

Few decades ago, trust was a feeling and a reality. The curiosity of 'feelings often diverging from reality' made trust a major constituent of social studies. Person A may believe that B is trustworthy, although that is not the case, but A may also believe in the opposite (B being non-trustworthy), although it may not be the case. This led to the creation of social models of trust. Nowadays, trust has also come to represent people's beliefs in the entities met in the virtual world, leading to the design of many computational models of trust. However, the models that represent the relationship between A and B can easily fail to capture both sides of the story: how much A trusts B, and whether B is really trustworthy. In other words, how much the model resembles reality, how much it comes closer to human perceptions and actions, and how much the two differ. In this study, we take the stance that trust has its own representatives in the online environment. We recognize reputation, rating, and recommender systems (henceforth denoted as RRR systems) as the online representative of trust. We see the elements of the three (RRR) systems as ones that are different, but in often complement one another. However, the adequate combination of RRR depends on many factors, and the failure to recognize those factors leads to inconsistencies in the

work of the RRR system as a whole. Moreover, user behavior is largely influenced by the work of those systems, and it largely influences the systems performance as well.

The contribution of our work is in the following: we detect and analyze four factors that influence the work of RRR systems when they are required to co-evolve as a single solution for providing good user experience, system reliability and result accuracy. Those factors are context, risk, dynamics and privacy. We put users in the core of RRR systems, and investigate the inter-dependencies between RRR systems and user behavior with respect to the four identified factors. We analyze several types of cognitive biases that users exhibit in their online experiences with RRR systems. Then, we perform an experimental study that will allow us to identify several types of cognitive biases exhibited by the users, and that were not investigated in an RRR setting so far. Based on the identified factors and the findings of the experimental study, we propose a framework for addressing some of the issues attributed to users' biased behavior and explore the possibility of employing hidden signals in the RRR systems for the purpose of capturing some types of user behavior.

To provide a clear picture of our understating of social trust and its online representatives, in the following section we define the concepts. Then, we present related work in the area with the effort to move closer to each other the social and the technical aspect. Finally, we present the stated contributions throughout Sections 4 and 5. We conclude in Section 6.

2 The Notion of Trust and its Online Representatives

2.1 Trust

From a social perspective, trust can be defined from two general aspects: cognitive and affective. The former is represented by concepts like rational choices, learning loops, institutional protocols, pattern detection, and imitation of established norms. Affective aspects, on the other hand, are mainly seen in the emotional side of trust interactions, and they account for the human feelings. As feelings are heavier on energy, whereas thoughts are heavier in information, affective properties often "take the blame" for contributing to cognitively biased decisions [1]. The 'social' literature on trust and reputation is exhaustive [1–3]. Clearly, when seeing trust as a purely social phenomenon, it can only be ascribed to living beings, and manifested through the property of trustworthiness of entities that are not necessarily living beings themselves. However, the ability to trust nowadays is disentangled from purely social contexts. A trusting entity can be any agent capable to resolve cognitive conflicts, or do preferential filtering, ranking, or sorting. This makes the definition and even the purpose of trust hard to grasp and determine. The blurred line of where the human factor starts or stops to influence trust and trustworthiness, usually leads to neglecting the affective side when defining the computational analog of trust. This, in turn, leads to the inability to predict the behavior of systems where trust models are deployed. These effects are caused by the highly non-linear nature of the trust phenomena, which do not allow a system to be designed according to the elegant principles of mathematical linearity and probabilistic averaging. Hence, knowing the composition of the system parts does not contribute a lot to inferring the properties of the system

as a whole. It is critical to consider the interactions and dependencies between the entities that comprise the system, and capture the additional phenomena and properties that emerge from those interactions. Complementing this with the strong contextual dependence of trust explains why researchers have a hard time formalizing trust, and incorporating it into online scenarios analogous to those in the traditional social networks. Following Gambetta [4], we give the following initial definition:

Definition 1a. Trust is the belief, i.e., the subjective probability that an entity will perform in a way likely to bring the expected benefit, or not do an unexpected harm.

Considering trust only as a *subjective probability* leaves out an extremely important concept related to trust, that of *risk*. This fact has also been the catalyst of a vigorous debate between economists and social psychologists [3]. In circumstances where one entity *relies* on another, trust choices include a certain level of risk. Josang [5] defines two different types of trust, *Reliability* and *Decision trust*. The former covers the aspect of trust as stated by Definition 1a. The latter considers the risk brought about by the uncertainty of transactional outcomes and is used to extend our first definition:

Definition 1b. Trust is the extent to which one entity is willing to depend on others' decisions and actions, accepting the risk of a negative (undesired) outcome.

2.2 Online representatives of trust

Despite the relatively interchangeable use of trust and reputation in the research community, it is essential to understand the difference between the two.

Definition 2a. Reputation is the general standing of the community about an entity's trustworthiness, based on the past behavior, performance, or quality of service of that entity, in a specific context, i.e., a domain of interest.

Definition 2b. A system that facilitates the process of calculating and managing reputation is called a reputation system.

Hence, reputation is the amount of context-aware trust, i.e., a quantitative representation of the trust that the society places in an entity, bound by the domain of interest. In addition to reputation systems, we consider rating and recommendation systems to also be online representatives of trust. We define them as follows:

Definition 3. Rating systems manage the evaluation or assessment of something, in terms of quality, quantity, or a combination of both.

Definition 4. Recommender systems are a subclass of information filtering systems that seek to predict the rating or preference that a user would give to an item or a social element they had not yet considered, using a model built from the features of an item (content-based) or the user's social environment (collaborative filtering) [6].

We use the terms *trust mechanisms* and *RRR systems* interchangeably in this work.

3 Related work

3.1 Social

Some of the work done in social and behavioral sciences that inspired computational trust research was discussed in the previous section [1–3]. Neuroscience has also re-

vealed that emotions and cognition, present in different areas of the human brain, interfere with each other in decision making, often resulting in a primacy of emotions over reasoning [7]. A very similar, although deceptively simple idea stands behind the outstanding work in Perceptual Control Theory: our perceptions are the only reality we can know, and the purpose of all our actions is to control the state of this perceived world [8]. The psychology of making trust-related choices is directly related to how people think and feel, perceive and decide. The brain has developed complex mechanisms for dealing with threats and risks. Understanding how it works and when it fails is critical to understanding the causal loop between trust-related perceptions and trust-related choices. An area with remarkable results about the irrationality, bias, and unpredictability of human actions in various circumstances and mindsets is Behavioral Economy [9–11]. In the context of preferential reasoning, their analyses show that users are often unaware of their taste, even for experiences from previously felt outcomes. Not only this reveals that taste is much more subtle than preference, but it shows that preference itself is not a stable property of human reasoning [12]. Experiments on persistency of user preferences about identical items at different instances of time proved significant fluctuation in the repeated preferential choices [10][13]. In contract and utility theory, the potential of employing trust mechanisms for dealing with information asymmetry was recognized long ago. When the possibility of post-contractual opportunism creates a context of moral hazard, trust mechanisms are employed for sanctioning undesired behavior. Another context of information asymmetry is adverse selection, and arises when one is required to choose a transaction partner whose type (good or bad) is unknown. In his work, Akerlof analyzes the effect of social and trading reputation on transaction outcome and market maintenance [14]. The study demonstrates that goods with low quality can squeeze out those of high quality because of the information asymmetry present in the buyers' and sellers' knowledge about the products – the problem of so called "lemon markets". Reputation mechanisms would balance this asymmetry, helping buyers make better-informed decisions by signaling the behavioral types of sellers, and at the same time providing incentives for sellers to exchange high-quality goods. Thus, Akerlof makes an instructive distinction between the *signaling* and the *sanctioning* role of RMs, which was only recently considered in computer science.

3.2 Technological

Understanding the behavioral implications of users in the field of computational trust is crucial, as the user factor in the processes met in online trust mechanisms is omnipresent. As online representatives of trust, RRR systems are also assigned the role of "devices" that help decision-making under information asymmetry (reputation and rating systems), and information overload (recommendation systems). In the former case, they have sanctioning and signaling role, and in the later case – directing and filtering role. The goal of employing RRR systems is to reduce the complexity that arises from information overload, and to lower the uncertainty present in the contexts of information asymmetry. Depending on the general context, however, the combination of the three (reputation, recommendation, and rating) requires different structuring to achieve the desired goals. This is discussed further in the following section. RRR systems rely to a great extent on preference inputs from users. Bias in these

inputs may have a cascading error effect on the performance of the applied algorithms. This does not only affect the accuracy of the results, but it also influences the perceived system reliability. Hence, user preferences are malleable and affect system performance, but they are also largely influenced by the information provided by RRR systems. Yet, biased behavior, its causes and effects, are relatively unexplored issues in the field. But the fact that only a narrow set of cognitive biases has been tackled by the research community does not imply there are no significant studies made in this regard. In [18] and [19], authors investigate the so called self-selection bias, whereby users only rate the items (movies) they like most, causing extremely high average rating for the rated items. Such ratings are representative only for a specific group of users, but do not truthfully depicture the item's general quality. Furthermore, self-selection bias was proved to not only be a transient phenomenon, but also a steady state in the system [15]. Although seemingly absurd, there is also one positive implication from this result of "self-selection bias" sustainability: if there was a strategy in the managerial principles of some company to cause inflated ratings for certain products by the "first-mover effect" [17], this effect would be flattened out on the long run. Such issues may appear to have purely economic nature, but they seriously compromise the reliability and performance of current RRR systems. The high impact of online reviews on product sales was demonstrated in [18], uncovering some of the motifs behind companies' efforts to appear competitive on the market. Positive rating bias was noted throughout systems of different nature. eBay is claimed to owe its success to its reputation/feedback system, yet out of the 57% of the users that decide to leave feedback, 99% issue positive feedback [19]. Moreover, large amount (41%) of users prefers to stay silent rather than to leave a negative feedback. A proposal to interpret silence as part of the user feedback was proposed in [20].

4   Cognitive Bias and its loops of causality in RRR systems

Cognitive bias describes a replicable pattern in perceptual distortion, inaccurate judgment, or illogical interpretation [10]. Clearly, bias can be noticed in both perceptions and actions, and the two are bound by the way people process their perceptions in order to take an action (including non-action). The paradox that arises in RRR systems is that, although user preferences are overly biased and affect the system performance and reliability, the preferences themselves are largely influenced by the ratings and the results provided by RRR systems. In this section, we explore the causes and effects related to cognitive bias in RRR systems, and we identify four major factors that influence these causal loops of biased behavior.

4.1   Context and Bias

Context is the set of circumstances or facts that surround a particular event or situation. Here, we analyze three general contexts in which an online interaction can take place: one with *pure collaborative* elements, one with *collaborative and competitive* elements, and one with *collaborative, competitive, and monetary* elements. Clearly, each subsequent context includes the elements of the previous, implying added system complexity. Therefore, it is crucial to understand which elements can be ade-

quately combined in order to meet a system design-goal, without encumbering its performance and flexibility to the extent of edging users out.

Once a reputation score becomes part of users' profiles, the users themselves become identifiable by their reputation, as if it guards their 'online brand'. Hence, maintaining a stable identity is crucial for the joined reputation value to make sense. On the other hand, being equipped with an online identity as a synonym for one's reputation gives rise to some new dimensions in the context of impression management. Once reputation is used as a signal for the user behavioral type, pure collaborative context ceases to exist. The requirement for conducting a successful impression management adds a competitive component to the system, and makes the reputation explicitly recognized as part of a person's social capital. Coupling this with the presence of bias implies that inflated (overly positive) reputation values in a system devaluates the reputation itself, as if the presence of reputation value defeated its purpose. Such a situation creates the need for additional incentives that will shift the reference for a good behavior from impression management to another context. What is often done in this regard is applying monetary elements to provide incentives for a desired behavior, which brings its own issues. Mixing purely collaborative context with monetary elements was already proved to throw a shadow on both the social intentions and the opportunity for monetary gain [10]. In addition to these more subtle influences, there are more detrimental effects that arise from inadequately combining such context-elements. Including monetary elements drastically increases the complexity of the system, and introduces inter-locking dependencies between the trust mechanisms and the outside environment. From a systemic perspective, this implies that the boundaries of the system are open to additional disturbances [21]. As a result, claiming predictability of the users' reputation scores, or moreover, of user and system behavior, diminishes, and failing to recognize this leads to system degradation, and eventually, system failure. This is also the core idea behind the Tragedy of Commons [22].

The soundness of the matters elaborated above was also demonstrated in practice. By announcing its Partner's Program[1] in May 2007, YouTube explicitly offered its highest rated users to earn revenue from advertisements placed next to their videos. This instantly triggered series of events of users blaming each other for using automated programs to inflate their videos' ratings[2]. While the effect of these blames is related to the tragedy of commons [22], the effect of using programmed agents to inflate one's own rating is known as the Cobra effect [23]: the solution of a problem makes the problem worse. These effects are often result of systemic ignorance, and only retrospectively analyzed by many system designers.

Pure collaborative contexts exist when a reputation is used internally in the system (for e.g., to provide a reference or serve as a regulator for the flow of some system processes), also implying keeping users' reputation private, or if the acquiring of reputation is not bound to one's performance, i.e. it is not used for signaling purposes.

Clearly, when making the decision of which contextual elements to choose as part of a system, context is intimately related to risk. With the addition of each element, the complexity increases, and the perceived risk is hardened by additional factors. Furthermore, privacy appears worthy of consideration as an option for limiting the

detrimental effects of added complexity. The goal of this study is to have a holistic look on RRR systems through the defined factors, rather than analyzing each of them independently. Therefore, the next section examines the link between risk and bias.

4.2 Risk and Bias

Risk is conceived as the possibility of triggering unexpected, unlikely, and detrimental consequences by means of a decision attributable to a decision maker [21]. Uncertainty is part of every online interaction. The extent of uncertainty, the expected utility, and the cost required for performing an action, influence the perceived risk a transaction brings. The field of prospect theory offers an incredible amount of experimental work demonstrating the myriad of cognitive biases that people exhibit when faced with risk and uncertainty [11]. A phenomenon that binds risk and uncertainty is the so called *pseudo-certainty effect*, which reflects the tendency of people to be risk-averse if the expected outcome is positive, but risk-seeking when they want to avoid negative outcomes. While it is a curiosity to inspect the properties of each bias independently, in reality biases are often coupled together, acting as both the cause and the effect of human perceptions and actions. The property of non-linearity we assigned to trust systems implies that complementing few biases together does not mean that their causes and effects will work in an additive fashion. Prospect theory has demonstrated that people underutilize consensus information, and when given descriptive facts about the quality of a person, they make choices regardless of the statistics offered about that person [11]. Josang et al. provide a formal proof of this phenomenon known as the *base-rate fallacy*, and give a formal framework for accounting for this fallacy in a computational setting [24]. Information offered in RRR systems is to a great extent a statistic produced by the collective efforts of the community members. This information can be represented in various ways - numerically, descriptively, as a single or multi-valued component, or as a combination of those. Therefore, it is important to explore how users perceive different types of information, and whether the descriptive and the numerical representation of an entity's quality can collide in users' perception. This is something we also investigate in our experimental study.

4.3 Dynamics and Bias

To some extent, we already touched on the issue of dynamics and bias when discussing the inconsistency of user preferences over time [10]. Here, we stay more in the context of trust and reputation, and connect the dynamics-factor with the previously defined – risk and context.

From the definitions of trust and reputation (Definitions 1 and 2), it becomes clear that the dynamics of trust differs from the dynamics of reputation. This discrepancy cannot be captured by any model, as both trust and reputation are in reality intangible matters. Yet, the social models and protocols for detecting malice seem to be successful. In RRR systems, one way the dynamics of reputation is embedded in the models is through discounting the relevance of gathered information by some time-factor. However, such approach disregards the importance that some information had in the past in terms of its impact on the outcome. In other words, discounting by recency and frequency is not equal to discounting by impact. Doing the former may provide disin-

centives for the users who take important actions at a lower rate. Closely related to this issue is the bias of *rosy retrospection*, which is a tendency to rate past events more positively than they were actually rated when the event occurred [11]. This reflects the importance of accounting for the time-interval between item-consumption and provided feedback about that item. Unfortunately, our current experimental study does not tackle any of the concepts related to dynamics. However, as a disintegrable part of the bigger picture of online trust mechanisms, dynamics must be taken into consideration. Our future work will devote more attention to this factor.

4.4   Privacy and Bias

In [25], authors show that individual interpretations of trust and friendship vary, and the two concepts are correlated to other characteristics of a social tie and to each other in a non-symmetric way. Furthermore, they provide evidence that raters consider how a ratee's reputation might be affected by the feedback. This fear of bad-mouthing and reciprocation in the context of impression management is directly related to the fear of retaliation in e-commerce systems [19]. Together, these factors also help to explain patterns like higher reciprocity in public ratings, and the near absence of negative ratings. Given this reluctance to publicly leave negative feedback, the question arises: why offer multi-valued choice for item evaluation in the first place. Moreover, why showing it publicly if it affects the users' decisions to an extent that makes it useless.

Closely related to privacy and the context of impression management are the concepts of individual and group behavior, and similarly, individual and group bias. Whereas most of the biases we mentioned so far were characteristic for an individual, we would fall extremely short on a useful discussion if we do not touch upon group behavior as well. After all, building reputation is essentially a social process, regardless of the fact that trust individuals cherish for one another underlies this process.

In [26], authors study the anchoring effect that item ratings have on user preferences. They find that users' inclination towards providing positive feedback is additionally amplified if users see the current rating that an item got by the rest of the community members. That people imitate, or do what others do, especially when having no determined preferences, is nothing new. This is often the cause of what is known as group polarization, bandwagon effect, or herd behavior, depending on the field of study that identifies it [9]. In the context of RRR systems, a study on group polarization on Twitter showed that, like-minded individuals strengthen group identity [27]. In other words, when part of group situations, people make decisions and form opinions that are extreme than when they are in individual situations. For RRR systems, this implies that it is not only important to acquire a significant amount of user feedback, but to also investigate whether this amount of user opinions was inferred from sufficient number of independent sources. A formal apparatus for resolving such issues in a computational setting can be found in [28], where the author provides a framework for reasoning with competing hypothesis using Subjective Logic.

The following section describes the experimental study and analyzes the findings through the factors defined above. Moreover, it proposes some ways to address the revealed issues.

5   Experimental work

Objectives: The major objective of this survey was to provide data that will help us investigate the compatibility between users' perceptions of the RRR system and its design objectives, but also reveal new directions for reasoning about the inter-relations between users and systems. The main questions we aim to answer are:

— Which descriptive model resembles closest user perception of numerical feedback?
— How does a slight difference in the 'tone' of presented choices in the two descriptive models influence users' decision?
— How does the presence/absence of different contextual elements influence user choice, and is it related to the nature (numerical or descriptive) of the alternatives?

Design and Methodology: Two types of methods were used to gather the necessary data – an online survey method[3] was chosen for better geographical spread of respondents, speed of data collection and independence of participants' opinions; and direct (one-on-one) interviews were performed to capture some of the subtleties that are ungraspable by only observing the outcomes of the systems. Such subtleties are the discrepancy between preferences of the majority of individuals and the group preference, the huge difference in the choices made as new contextual elements are introduced or taken away, the difference and inter-dependence between trustworthiness and acquaintanceship, etc. Each survey contained the same two questions, but offered slightly different evaluation choices. The questions represented real reviews for HP Laptop taken from *Epinions*[4]. To not disturb the flow of the paper, they are given in the Appendix. All three groups were asked to rate the reviews for their usefulness. The first group (Survey 1) was asked to give a numerical rating on a scale 1 to 5 (1 = lowest and 5 = highest rating). The second and the third group were offered descriptive evaluation choices. However, among the possible answering choices for Survey 2, there were also such that stated explicit negative experience (Not useful at all), whereas the answering choices for Survey 3 varied from neutral to positive.

Table 1. Statistics about the experimental setting

Survey Info. Respond. Info.	Type of Survey			
	Survey 1	Survey 2	Survey 3	
	1-5 Num. Feedback	Neg.Descriptive	Neutral-Pos. Descriptive	Total
Answering choices	1 2 3 4 5	Not useful at all Hardly useful Somewhat useful Quite useful Extremely useful	Neutral Somewhat useful Quite useful Very useful Extremely useful	
Responded	22	22	22	66

Respondents: The experimental work is conducted over a population (both female and male) of 86 people, at the age of 20 – 50. Its completion required no special technical knowledge, and subjects had no difficulty understanding the assignment. Respondents were divided into 5 groups. Three groups of 22 people were formed for each survey type. Table 1 summarizes these statistics. 30 of the 66 survey respondents

---

[3] www.surveymonkey.com
[4] www.epinions.com

were also additionally interviewed. The results from these interviews are presented in the final subsection. In addition to the three groups, another 20 users were asked to independently evaluate each review (10 users per review). They expressed their evaluations both numerically, as in Survey 1, and descriptively, as in Survey 2.

Results: The following subsections show the major findings from this experimental study. Although the experiment was of relatively small scale, some interesting results were revealed with respect to the given objectives.

5.1   Distinction bias

As a first step in revealing the potential presence of cognitive bias in our experimental setting, we compare the results from the individually evaluated reviews with those obtained by the three surveys, where the two reviews were put together for evaluation. The goal is to reveal the potential presence of a *distinction bias*, manifested as a tendency to view two options as more dissimilar when evaluating them simultaneously than when evaluating them separately. This bias is often exploited in commerce scenarios, when sellers aim to sell a certain product (anchor) by placing it along with another - decoy product that appears as the worse option when put together with the anchor [10]. Fig.1 presents the results from the 20 independent evaluations of the two reviews. The horizontal axis shows the rating values, whereas the vertical – the number of users who provided the rating. As shown, both reviews were evaluated as equally useful, 2 on a 1 to 5 scale, and descriptively qualified as Somewhat Useful.
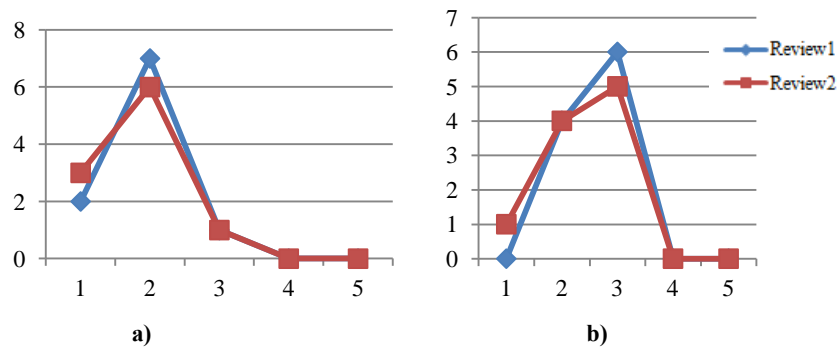

Fig. 1. Rating distribution for each review by offline users a) Numerically b) Descriptively

Fig.2 shows the difference in ratings between the two reviews for each of the surveys. Clearly, that they do not match with those shown in Fig1. In all three cases, Review 1 was evaluated as being better than Review 2. The mean values for the ratings of Review 1 and Review 2 for the three surveys are given in Table 2. Since preferences are often formed through distinction between given options, joint evaluation of recommendations may often result in a choice mismatch. The consequence is that the choice that appeared as the best option may not provide the best user experience, leading to dissatisfaction. Clearly, the issues of distinction bias and anchoring effect are important to account for in the design of RRR systems. The question is: what is the cause of those biases, what their effects are, and how to account for them in practice. The next section aims to provide the answer.

Table 2. Mean value for the ratings of the surveys

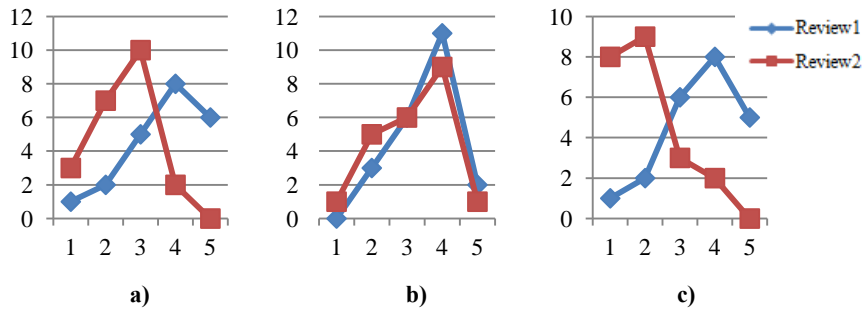|	1-5 Num. Feedback	Neg. Descriptive	Neutral/Pos. Descriptive
Review 1	3,727	3,545	3,636
Review 2	2,5	3,18	1,95

Fig. 2. Rating Distribution for the two reviews for a) Survey 1, b) Survey 2 and c) Survey 3

5.2 Numerical-Descriptive discrepancy and Positive Bias

In order to address distinction bias and anchoring effect, we must first understand if and how the presented evaluation choices affect the user opinion. This section investigates which of the descriptive surveys comes closest to the one with numerical ratings. The practical implication is in exploring if the current RRR systems that offer numerical ratings really match the user understanding about the meaning of those ratings. For that purpose, we compare the rating distributions of the two reviews for Survey 1 to those of Survey 2 and Survey 3.
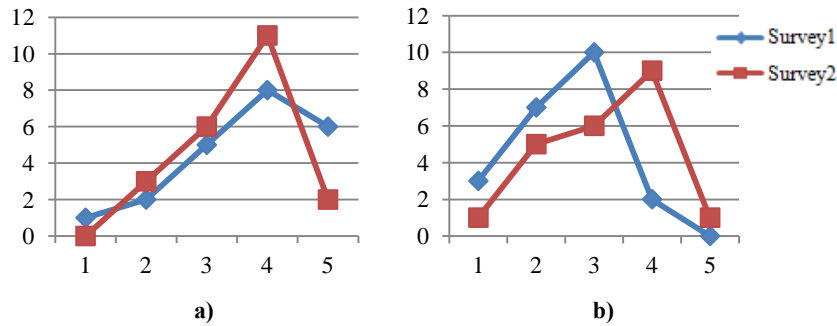

Fig. 3. Difference in Rating Distribution between Survey 1 and 2 for a) Review1 b) Review2

The results are shown in Fig.3 and Fig.4 respectively. It can be noticed that there is a good match for the distributions of Review 1 in both of the cases. However, compared to the distributions on Fig.1, there is still a great difference between the average rating value for Review 1 (1.9) provided by the independent user evaluations, and the average rating value for Review 1 in Survey 1 (3.727), Survey 2 (3.545) and Survey 3 (3.636). We identify the following causal link: distinction bias is caused when Review 1 and Review 2 are put together for evaluation, whereby Review 1 appears as the

better option; moreover, Review 1 also appears as one of exceptional quality, leading to a positive bias and exaggerated positive rating.
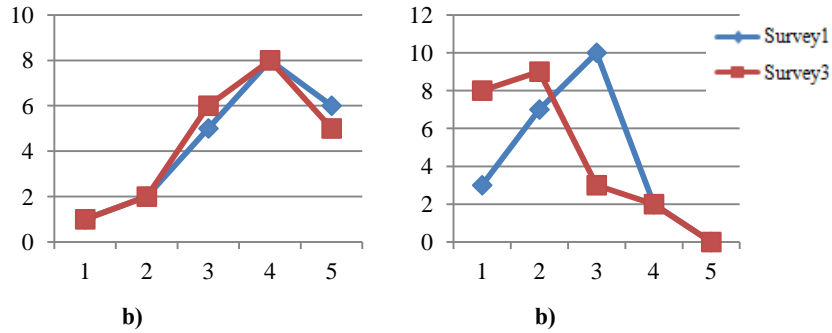

Fig. 4. Difference in Rating Distribution between Survey 1 and 3 for a) Review1 b) Review2

5.3   Positive bias and the Framing effect

In this section, we analyze the two surveys with descriptive evaluation choices. They differ slightly by the tone of positivism, although both offer five choices. Survey 2 contains more explicit negative statements, while the choices in Survey 3 vary from neutral to positive. The question we want to answer is *How does a difference in descriptive choices affect users evaluation*? To do that, we compare the rating distribution of the reviews for Survey 2 and Survey 3. The results are given in Fig.5a) and b).
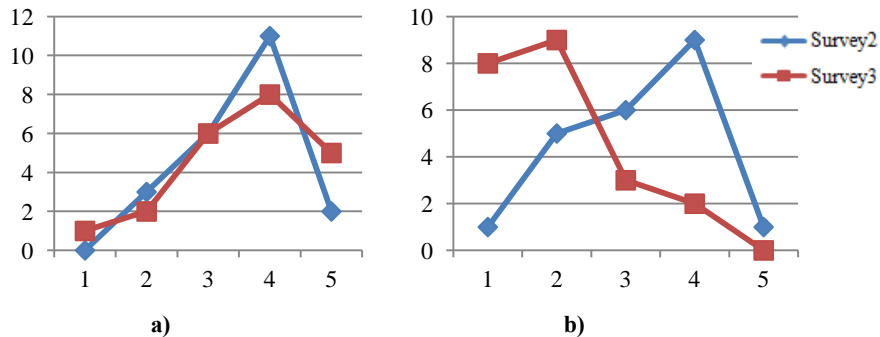

Fig. 5. Difference in Rating Distribution between Survey 2 and 3 for a) Review 1; b) Review 2

They reveal large difference in the evaluations of the reviews between the surveys, although the offered choices differ only slightly. This demonstrates that people tend to draw different conclusions from the same information, depending on how that information is presented, known as the *framing effect*. On Fig.5a), this effect is demonstrated as a slight smoothing of the exaggerated positives. Complementing these results with those shown on Fig.2b), we can conclude that for Survey 2, the ratings for the two reviews also come close to each other. One interesting result is the high mean rating value for Review 2 in Survey 2. Compared to the results on Fig.1, we see that the presence of a positive bias is higher when users are offered to chose between neg-

ative-positive evaluations, compared to a setting where they are required to choose between neutral-positive. In the next section, we use these findings to form our proposal on how to address some of the issues of the explored biased behavior.

5.4 Proposal: Hidden signals and "shades of grades"

The remarkably higher number of neutral evaluations demonstrated by the experiment (Fig.5b)), which is even higher than the positive evaluations for Review 2 (Fig.5a)), not only confirms the reluctance of users to give negative ratings, but also points out the importance of neutral vote as a connector between negative and positive. In addition to introducing neutral as a "shade" between negative and slightly positive, another finding in our study is that introducing Very Useful as a shade between Quite Useful and Extremely Useful shifts the exaggerated positives towards lower ratings. Both of these effects smooth the effects from positive bias, but also better capture people's perception of the offered choices. This might also explain why the offline model of social reputation succeeds to detect untrustworthy individuals without requiring consensus on someone's trustworthiness. The offline world offers numerous opportunities to pick on the hidden clues behind people's intentions. We consider the idea of introducing such hidden signals in the online RRR systems worth exploring.

The framework we propose for addressing the presented issues of biased behavior can be summarized as follows: first, by accounting for the specific context elements, we propose disentangling the collaborative, competitive and the monetary elements when deciding on the RRR design. Second, by accounting for those context elements, a decision should be made about which of them are desirable as public features. Then, it is crucial to determine the right representation of publicly displayed features in a way that fits the users' perception of the feature's meaning. Finally, by introducing hidden signals and shades of grades about the qualitative types of the entities, a better and more truthful distribution of the results can be obtained.

5.5 Closing the loop: The market of lemons on the market of opinions?

This section will close the loop of our study on cognitive bias by returning to the point where we started the discussion – context. The impact of context in the formation of cognitive bias required more interactive work with the respondents of the survey. Therefore, we additionally interviewed 30 people. The results are the following:
Q1. If you know that, depending on the ratings given for their reviews, reviewers will get proportionally higher/lower amount of money, would you give the same grade?
    22 answered: *No, one or two grades lower*. **1** was *Not Sure*; **7** answered *Yes*;
Q2. What if you yourself were a reviewer and the amount of money you would get depended on the amount of money other reviewers for the same product get?
    27 answered *Definitely a lower grade*; **3** answered *Still the same grade*.
The 3 who gave the 2b) answer were additionally asked:
Q3. If you see that the opinion you consider of low quality is the one that got the highest number of votes, would you reason the same the next time?
    2 answered: *No*; **1** answered *Yes*.

The purpose of these questions was to investigate the reasoning of the respondents as they were required to switch between contexts. The interview is of very small

scale, but still pointed to new directions for reasoning about user behavior through the elaborated contextual elements. There is, however, a deeper meaning of the obtained results: the exaggerated positives, the slight change in evaluation options followed by high difference in evaluations, and the rest of the biases we explored, may be merely the effects of user behavior in RRR systems. Different combination of particular contextual elements leads to different manifestation of the effects from those biases. Introducing monetary elements must be done with great caution, as it may cause information to be treated as a limited resource of monetary value, or as a trading resource in the process of acquiring social capital. This in turn leads to squeezing entities (items, agents, users) of potentially high quality out of the system, leading to what we refer here to as "the market of lemons on the market of opinions".

6   Concluding Remarks and the Way Forward

Trust is a feeling, a model, and a reality, with people in their core. Understanding how they work and how closely they resemble each other is essential for their design practices. Our work detected and analyzed context, risk, dynamics and privacy as factors that influence both the work of RRR systems and the users' understanding of that work. We explored the relation between RRR systems and user behavior with respect to those factors and analyzed few types of biases exhibited through the users' online experiences with RRR systems. In an experimental study that included 86 users, we found that users exhibit distinction bias, positive bias, anchoring effect, and framing effect. These have not been investigated in such a holistic manner for any of the current RRR systems. Based on the identified factors and the findings of the experimental study, we proposed a framework for tackling some of the issues attributed to users' biased behavior and address the possibility of employing hidden signals and shades of grades in the RRR systems for the purpose of capturing some of the detected biases.

Our future work will concentrate on investigating more the factor of dynamics. As we already referenced Perceptual control theory and Subjective Logic as an apparatus that provides formal reasoning with respect to human beliefs and perceptual behavior, the major part of our work will be directed towards joining the ideas of the two and employing them to formalize trust relationships under biased user behavior.

References

1. R. B. Zajonc, 'Feeling and thinking: Preferences need no inferences', *American Psychologist*, vol. 35, no. 2, pp. 151-175, Feb. 1980.
2. John Conlisk, 'Why Bounded Rationality?', *Journal of Economic Literature*, vol. 34, no. 2, pp. 669-700, 1996.
3. C. Castelfranchi and R. Falcone, 'Trust is much more than subjective probability: Mental components and sources of trust', *32nd Hawaii International Conference on System Sciences - Mini-Track on Software Agents, Maui*, vol. 6, 2000.

4. D. Gambetta,'Can We Trust Trust?'*Trust:Making and Breaking Cooperative Relations*

5. A. Josang, R. Ismail, and C. Boyd, 'A survey of trust and reputation systems for online service provision', *Decision Support Systems*, vol. 43, no. 2, pp. 618-644, Mar. 2007.

6. F. Ricci, L. Rokach, and B. Shapira, 'Introduction to Recommender Systems Handbook', in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 1-35.

7. R. MacMullen, *Feelings in history, ancient and modern*. Regina Books, 2003.

8. W. T. Powers, *Behavior: The Control Of Perception*,2Benchmark Publications, Inc., 2005.

9. D. Kahneman, 'Maps of Bounded Rationality: Psychology for Behavioral Economics', *American Economic Review*, vol. 93, no. 5, pp. 1449-1475, 2003.

10. D. Ariely, *Predictably Irrational: The Hidden Forces That Shape Our Decisions*, 1st ed. HarperCollins, 2008.

11. D. Kahneman and A. Tversky, 'Subjective probability: A judgment of representativeness', *Cognitive Psychology*, vol. 3, no. 3, pp. 430-454, Jul. 1972.

12. D. Ariely, G. Loewenstein, and D. Prelec, 'Tom Sawyer and the construction of value', *Journal of Economic Behavior & Organization*, vol. 60, no. 1, pp. 1-10, May 2006.

13. D. Cosley, S. K. Lam, I. Albert, J. A. Konstan, and J. Riedl, 'Is seeing believing?: how recommender system interfaces affect users' opinions', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, 2003, pp. 585–592.

14. G. A. Akerlof, 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *The Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488-500, 1970.

15. M. Kramer, 'Self-Selection Bias in Reputation Systems', in *Trust Management*, vol. 238, Springer Boston, 2007, pp. 255-268.

16. X. Li and L. M. Hitt, 'Self-Selection and Information Role of Online Product Reviews', *Information Systems Research*, vol. 19, no. 4, pp. 456 -474, Dec. 2008.

17. M. B. Lieberman and D. B. Montgomery, 'First-mover advantages', *Strategic Management Journal*, vol. 9, no. S1, pp. 41-58, Jun. 1988.

18. J. A. Chevalier and D. Mayzlin, 'The Effect of Word of Mouth on Sales: Online Book Reviews', *Journal of Marketing Research*, vol. 43, no. 3, pp. 345-354, 2006.

19. P. Resnick and R. Zeckhauser, 'Trust among strangers in internet transactions: Empirical analysis of eBay' s reputation system', in *Advances in Applied Microeconomics*, vol. 11, Bingley: Emerald (MCB UP), pp. 127-157.

20. C. Dellarocas and C. A. Wood, 'The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias', *Management Science*, vol. 54, no. 3, pp. 460 -476, Mar. 2008.

21. N. Luhmann, *Risk: a sociological theory*. Transaction Publishers, 2005.

22. G.Hardin, Tragedy of the Commons,*Science*,vol.162,no.3859,pp.1243-1248,Dec. 1968.

23. H.Siebert,*DerKobra-Effekt* :wie man Irrwege der Wirtschaftspolitik vermeidet:Piper, 2003.

24. Audun Josang and S. O'Hara, 'The base rate fallacy in belief reasoning', in *2010 13th Conference on Information Fusion (FUSION)*, 2010, pp. 1-8.

25. L. Adamic, D. Lauterbach, C.-Y. Teng, and M. Ackerman, 'Rating Friends Without Making Enemies', *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pp. 1-8, 2011.

26. J. Zhang, 'Anchoring effects of recommender systems', in *Proceedings of the fifth ACM conference on Recommender systems*, New York, NY, USA, 2011, pp. 375–378.

27. S. Yardi and D. Boyd, 'Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter', *Bulletin of Science, Technology & Society*, vol.30,no.5,pp.316 -327,Oct. 2010.

28. S. Pope, 'Analysis of Competing Hypotheses using Subjective Logic', *Systems Technology*, no. June, pp. 13-16, 2005.

Appendix: Survey Questions (Reviews)

Review 1:
User Rating: OK; Ease of Use: 2;Quality of Tech Support: 1
Pros:1)Intel core i3 processor;2)Finger Print Sensor;3)Battery life
Cons:1)Build material;2)Intel integrated graphics;3)320GB hard disk space
The Bottom Line: This laptop is not recommended to any one, because of poor support quality. The products of HP were good 3 years ago. People loved them for their looks, reliability, performance and budgeted price. But after 2008 the reliability ratio of HP as compared to its competitors is far below in almost all aspects including price, performance, looks and the most important reliability. I myself was a big fan of this company but after facing the failure of consecutively two the PCs' from this company, I changed my view. My wife had this laptop and I bought this laptop for her as her birthday present three months ago, but just after two months, the screen got dead spot (small black spots) and the hing connecting the screen with keyboard also got broken. Still the laptop was running fine but last week its keyboard also stopped working, which sucks. All the money I paid for it gone in vain. My experience about: PROCESSOR: Intel core i3 processor clocked at 2.1 GHZ works really very fine for multitasking but it is not designed to handle more tasks. This pc is fine for web-surfing, word-processing, office work, watching movies. GRAPHICS: Intel integrated graphics are not capable of running blue ray movies silky smooth. BUILD MATERIAL: This laptop is made of seriously very cheap plastic. Its glossy surface is just a fingerprint magnet and you have to clean up the laptop after every single use. The glossy screen also causes panic while watching movies in sunlight. LACK OF HDMI PORT: This laptop lacks some of the most important port which is included in the laptops of this price range, i-e; HDMI port. The sound quality is good, battery life is also impressive and last about 3 hours (6 cell) even when watching movies. The security feature like Finger Print Sensor works great. Its light weight makes it easier to carry this laptop everywhere but it can't be called ultraportable laptop. Over all this laptop isn't recommended to any one whether the person is student or house wife.
[Recommended: No]

Review 2:
User Rating: Excellent; Ease of Use: 5; Quality of Tech Support: 4
Pros: I would buy another computer from HP
Cons: multiple hardware faults. (i.e hard to type and sensitive mouse pad)
HP ProBook 4530s is a extremely good product. The computer is lightweight and rugged. It's able to be put in a backpack and stuffed wherever you need to go. When i bought and received the computer i liked how the computer didn't come with a lot of extra software junk. Next, i like having a numeric pad and i will continue to buy computers with them from now on. For just web browsing and playing light computer games this computer is extremely fast. The key pad is not sensitive enough and i have to hit the keys hard a lot. Also if you have multiple fingers close to the mouse pad, the mouse pad goes crazy due to its many settings. The speaker system is excellent the loudest system i have heard on any computer system. The software for the finger ID gets weird every now and then i have to input my fingers a couple of times to log in. The settings are easy to change on the computer. The battery life is average to other computers. The computer hardware is nicely laid out and easy to find. Computer is worth the price i paid for it.
[Recommended: Yes]