![JASSS logo]

Adam Wierzbicki, Paulina Adamska, Katarzyna Abramczuk, Thanasis Papaioannou, Karl Aberer and Emilia Rejmund (2014)

Studying Web Content Credibility by Social Simulation

Abstract

The Internet has become an important source of information that significantly affects social, economical and political life. The content available in the Web is the basis for the operation of the digital economy. Moreover, Web content has become essential for many Web users that have to make decisions. Meanwhile, more and more often we encounter Web content of low credibility due to incorrect opinions, lack of knowledge, and, even worse, manipulation attempts for the benefit of the authors or content providers. While mechanisms for the support of credibility evaluation in practice have been proposed, little is known about their effectiveness, and about their influence on the global picture of Web content production and consumption. We use a game-theoretic model to analyze the impact of reputation on the evaluation of content credibility by consumers with varying expertise, in the presence of producers who have incentives to manipulate information.

Keywords:
Credibility, Reputation, Game Theory, Incentives, Online Communities

Introduction

1.1   The Internet has become an important source of information that significantly affects social, economic and political life (Rainie & Kommers 2002). The content available in the Web is the basis for the operation of the digital economy (Melnik & Alm 2002; Resnick et al. 2006). Moreover, Web content has become essential for many Web users that have to make decisions concerning shopping, employment, education, health, finance, investment, etc. (Senecal & Nantel 2004; Chevalier & Mayzlin 2006). For example, Web content is employed by individuals to choose the right investment or product, and even for the purpose of self-diagnosis of disease and treatment selection (Weng et al. 2013; Johnston, Crooks, & Snyder 2012). The credibility of the content, that is, the extent to which one can rely on information from the Web, is of crucial importance.

1.2   Meanwhile, more and more often, we encounter Web content of low credibility due to incorrect opinions, lack of knowledge, and, even worse, attempts of manipulation for the benefit of the authors or content providers (Thompson 2003; Dellarocas 2006). Selecting highly-credible Web information is difficult. Search engines try to take into account the credibility of the content in their results indirectly, using the hyperlink structure as means of verifying the quality and popularity of the content. However, this is often insufficient to assess the credibility of information, as any experienced Web user is well aware.

1.3   There exist a number of services available online that aim to assess the reliability of Web content. One can name services such as Digg or Reddit, thematic portals and forums, news sites, recommendation sites like Epinions, as well as Question and Answer type portals like Quora. Most of these services use mechanisms based on a selection done by users that may, in a manner specific for a given service, select popular, interesting or high quality content. There are also first attempts to create services that are supposed to assess credibility directly, such as myWot.com, the hypothes.is or Reconcile projects. Portal myWot.com allows its users an assessment of Web content using the following criteria: trust, privacy, safety for children and reliability of sellers (in the case of e-commerce sites). It also provides commercial sales of credibility certificates. The hypothes.is project is a crowd-funded start-up from Silicon Valley, which seeks to develop methods of reviewing Web content. The purpose of reviews would be the assessment of the credibility of the content. Reconcile is a scientific project that aims to create Web-based tools for the direct support of credibility evaluation[1].

1.4 While the mechanisms for the support of credibility evaluation in practice have been proposed, little is known about their effectiveness, and about their influence on the global picture of Web content production and consumption. *What are the critical factors that affect the effectiveness of credibility support? Can credibility support have a positive influence on the behavior of content producers, and under which circumstances? Is the reputation of a Web content provider an important factor in the selection of credible content by consumers?* Such research questions cannot be answered easily, because of the relative immaturity of credibility support mechanisms on the Web, the lack of available data, and the difficulty in conducting large-scale experiments on the Web. However, we believe that some of these questions can be studied using social simulation. Social simulation becomes especially important for the study of what-if scenarios that concern the behavior of Web content producers and consumers and the impact of this behavior on effectiveness of the credibility support.

1.5 The first difficulty in approaching this problem is the lack of available models. While it would be possible to model a market of Web content production and consumption directly, using simulation, such model could become very complex. In an attempt to create a simple model that would capture the salient characteristics of Web content credibility evaluation, we have created the Credibility Game (Abramczuk et al. 2012). In its simplest form, this game-theoretic model can be analyzed mathematically. It is also very good basis for a simulation model that can be used to study more complex behavior of the two types of agents: Web content producers (CP) and consumers (CC). In this paper, we base on our previous work on the Credibility Game and extend our study to incorporate a fundamental mechanism used for Web content evaluation and selection: reputation.

1.6 We then use a simulation approach, developing a simulator of the Credibility Game with reputation that is used to verify the following hypotheses:

1. Consumer expertise constitutes a critical factor that affects the choice of producers whether to produce truthful information or to invest in presentation. Dishonest producers take advantage of the absence of reputation and of low consumer expertise to produce false information.
2. When reputation is preserved (no whitewashing), consumers use reputation to complement their assessment as a way to evaluate credibility of information.
3. Reputation (in absence of whitewashing) that is based on the actual consumer experiences leads to better results for the community (higher profits for the community members) than reputation that is based on superficial evaluation of Web content.

1.7 Apart from the verification of the above hypotheses, the main contribution of this paper is a simulation model of Web content production and consumption that can be used in studying Web content quality, credibility, and Web content selection mechanisms. Among the interesting features of this model, there is an explicit modeling of properties of information that are relevant to quality or credibility, such as presentation of information and source reputation. The model also allows to study subtle effects such as the difference between source reputation that is based on a first impression (i.e., imprecise signal that is a function of information presentation and truthfulness) and source reputation that is based on experience (i.e., consequences of relying on the consumed information). We also show how the proposed model can be extended to the model Web2.0 services such as Question and Answer sites. The proposed model can easily be used in studying adversary strategies against mechanisms of Web content selection or credibility evaluation support, but this is left for future work.

1.8 The rest of the paper is organized as follows. In the next section, we describe related work. In section Credibility Game Model, we describe the model in its basic version, its extension to a signaling game (Abramczuk et al. 2012) and the extension of the model with a simple reputation system. Section Simulator Design describes the used simulator. In section Impact of Consumer Expertise on Credibility Evaluation we describe the simulation results of the signaling Credibility Game that consider the effect of consumer expertise, modeled by variations of the received signal, on producer strategies. *In this section, hypothesis 1 is verified*. Section Impact of Reputation on Credibility Evaluation describes the main results of this paper, which are used to *verify hypotheses 2 and 3*. Section Conclusions and Future Work concludes the paper and describes our planned experiments and simulator extensions.

Related work

2.1 The problem of information asymmetry (i.e. hidden information at one of the transacted parties) has been extensively studied in economic theory. Hidden information may refer to a hidden quality (Wilson & Kreps 1982), and in that case adverse selection leads to the "market-of-lemons" effect, or to a hidden type, in which case moral hazard (Dellarocas 2005) deteriorates participation to the market. The problem of assessing Web credibility may involve both cases of hidden information, i.e. hidden quality for static pages or hidden type for dynamic pages. Reputation is considered as an adequate mechanism to deal with hidden information (Dellarocas 2005; Wilson & Kreps 1982). Such approach would necessitate the existence of credible ratings for the Web pages subject to assessment. Although initial attempts towards that goal have started (e.g. Google+, plus.google.com), it will be long before enough adequate user feedback will be available for the Web. Furthermore, easy domain name changes and dynamic content generation render associating historical information to a particular Web content rather difficult. To this end, in our work, we refrain from modeling the Web credibility problem as a reputation game and leave the in-depth investigation of this issue as a future work.

2.2 There is an extended literature on game models for online reputation systems (Dellarocas 2005; Papaioannou & Stamoulis 2008),

for a survey, see (Friedman, Resnick, & Sami 2007). These games study the incentives of transacted parties in an encounter towards honesty (Dellarocas 2005) and truthful feedback (Papaioannou & Stamoulis 2008). Guttman (1992) studies voluntary cooperation in case of labor-managed firms as a Prisoner's Dilemma (Chammah & Rapoport 1965). The problem of cooperation has two aspects: how to avoid free-riding and how individuals convince their fellow members that they will uphold their commitments. Cooperation can be an equilibrium outcome, when rational individuals act so as to preserve reputations for cooperating. The theory developed in this work implies that (a) voluntary cooperation will be negatively related to population turnover, and (b) the relationship of community size to voluntary cooperation will have an inverted-U shape. Instead, our work analyzes the incentives of Web content producers and consumers regarding the truthfulness of published information. To the best of our knowledge, there is no prior work that studied the problem of Web credibility as a game.

2.3   Jain, Chen and Parkes (2009) and McAfee and Ghosh (2011) address the issue of incentivizing high quality of user generated content in a game-theoretic modeling. In a more generic model, McAfee and Gosh (2011) study the economics of high-quality user generated content from strategic and exposure-motivated contributors. Their model predicts that if exposure is independent of quality – as it can be in our case – there will be a flood of low quality contributions in equilibrium. They propose a simple model for eliminating low-quality content, based on a single negative rating. However, such a model and approach are mostly relevant for blog comments and reviews, and could not be employed for arbitrary Web sites – such an approach could easily lead to defamation of honest Web sites by rational competitors.

2.4   Glazer and Rubinstein (2006) and Sher (2011) study a game of persuasion. In this game, a speaker attempts to persuade a listener to take an action by presenting evidence. The conditions under which the request is justified, from the listener's point of view, depend on the state of the world, which is known only to the speaker. Each state is characterized by a set of statements among which the speaker makes his choice. A persuasion rule specifies which statements the listener finds persuasive. Glazer and Rubinstein (2006) study persuasion rules that maximize the probability that the listener accepts the request if and only if it is justified, given that the speaker maximizes the probability that his request is accepted. They prove that there always exists a persuasion rule involving neither randomization (i.e. the listener can apply a random device to determine which hard evidence he asks the speaker to present) nor commitment (i.e. listener commits and announces an acceptance rule) and that all optimal persuasion rules are ex-post optimal. Sher (2011) shows that concavity is the critical assumption for both results: no value to commitment and no value to randomization. The key assumption is that the listener's utility function is a concave transformation of the speaker's utility function. Their model is broaden beyond binary listener's decisions given that concavity implies credibility. Assessing Web credibility somewhat resembles a persuasion game in the sense that the content producer (i.e. speaker) knows the credibility of the content (i.e. state of the world) and may use presentation quality as a persuasion rule for the content consumer (i.e. listener). However, in the Web credibility game the consumer cannot systematically reason and justify the information credibility given the producer's persuasion rules. Moreover, our signaling game model covers the case of consumers with different levels of expertise.

2.5   Regarding Web credibility assessment, Tanaka and Yamamoto (2011) identify six measurable factors related to the five main recognized factors (i.e., accuracy, authority, objectivity, currency, and coverage of topic) for judging the credibility of Web information, namely referential importance, social reputation, content typicality, topic coverage, freshness, and update frequency. On the other hand, Fogg et al. (2003) employ prominence-interpretation theory to explain credibility assessment; a site is deemed as credible by a user, if it contains features deemed as credible by that user. Other methods that automatically assess credibility aggregate different features values such as (Metzger 2007): information about credentials, advertisements, Web page design, type of Website, date of update, sentiment analysis, pre-defined search-engine page ranking, information commonality, source independence, source prestige, experience with the source and authority of information source. Our game model assumes that consumers assess content credibility based only on a signal regarding truthfulness and presentation quality passed from content producers. However, more features in the signal could also be considered.

🌐 Credibility Game Model

3.1   The Credibility Game is a game-theoretic model of Web content production and consumption that has been developed for the purposes of studying phenomena and behaviors related to Web content credibility, quality, and Web content selection mechanisms. The basis of the model is patterned after information theory: information is exchanged between a sender (producer) and receiver (consumer). Next, the model incorporates properties of the communicated information (metainformation) being the basis of payoffs, signals, and behavior. All these elements are included in a highly simplified manner in order to enable analysis. We also omit some important features of the situation such as e.g. information topic or information search.

Payoffs in the Basic Credibility Game

3.2   Basing on our previous research (Abramczuk et al. 2012), we assume that there exists a population of actors some of whom are assigned a role of content producers (CP) and some a role of content consumers (CC). These roles are fixed. The number of consumers is ten times larger than the number of producers, namely we have 1000 consumers and only 100 producers. It can be argued, how accurately this estimation reflects a real community, however it is difficult to verify the exact numbers in the Internet due to anonymity of certain actions. Our estimation is based on the assumption that there is an extremely limited group of users, who only produce information (for instance online newspapers), while most Internet users, although allowed to perform both consumer's and producer's actions, usually are not particularly active as producers. This inequality of contribution can be

observed in a microscale in certain communities (Ortega et al. 2008; Voss 2005). Taking into account the influence of number of producers on the overall reputation system's performance (smaller number of producers yields more interactions with consumers per producer, and therefore more evaluations) we have decided to increase the number of active producers to 10% in order to make it more difficult for the community to eliminate false information. In each round each consumer interacts with exactly one producer. The producer is chosen according to the current matching rule. In the basic model the matching rule is a simple random draw from the population of all producers. In an encounter, CP produces and CC consumes information.

3.3   Following the work on persuasion games, the basic choice of CC is binary: *CC can Accept (A) or Reject (R) the version of information produced by CP*. CP's strategies involve choosing the properties of the produced information *I*, namely its *truthfulness* (*TF*) and *look* (*L*), with $TF \in (0, 1)$ and $L \in (0, 1)$. Information truthfulness refers to a property of information that makes it reliable: information with *TF=1* can be completely relied on by the CC, which is reflected in a high payoff for the consumer who accepts this information. On the other hand, information with *TF=0* cannot be relied on, and the CC will receive a low payoff if he accepts this information. However, information with *TF=0* should not be thought of as a random information; rather, it is a plausible lie, or information that is in some parts correct, but contains an important error or omission that makes it costly to rely on.

3.4   Note also that *the TF information property can also be interpreted as a quality of information*, since this is a property that has a direct impact on CC's payoffs if CC accepts the information

3.5   Information look (*L*) refers to the presentation quality of the produced Web content. Increasing *L* can increases the likelihood that CC will accept the produced version of *I*.

3.6   Both truthfulness and look are binary variables in the simulation model. For this reason, *any content producer chooses one of four actions that represent the combinations of TF and L values* for the produced information *I*: Good looking Truth (*TF=1*, *L=1*); Bad looking Truth (*TF=1*, *L=0*); Good looking Falsehood (*TF=0*, *L=1*); Bad Looking Falsehood (*TF=0*, *L=0*). We shall refer to these actions as GT (Good looking Truth), BT (Bad looking Truth), GF (Good looking Falsehood) and BF (Bad looking Falsehood), for short.

3.7   The preferences of CP are modeled as follows. CP prefers that his produced information *I* is consumed by CC. Increasing look *L* of the produced information models an investment that CP can make. Yet, this investment is costly (increasing *L* increases the production cost), and if CC rejects the produced version, CP will incur a loss. On the other hand, if CC accepts, then the preference of CP depends on the TF property. We can have various types of producers with various preferences. An honest CP will prefer that CC would accept a truthful version of I: CP prefers that CC accepts I with a high *TF*. On the other hand, for a dishonest CP, the preferences may depend on a profit that can increase, if CC accepts a version of *I* that is less truthful. Consider, for example, a dishonest car salesperson that produces descriptions of cars on sale. CP's profit increases if CC buys a low-quality car that is described as a high-quality one. Therefore, lower *TF* yields a higher profit. This is a case of incentive incompatibility of CP and CC.

3.8   The consumer prefers to consume the information produced by CP if it has a high *TF*. On the other hand, the payoffs of CC do not depend on *L*, but only on *TF*. CC always prefers a content with a higher *TF*, and does not want to consume content with *TF=0*. CC prefers consuming *I* with *TF=1* to rejection, but prefers rejection to consuming *I* with *TF=0*.

3.9   For the basic simulation model we have chosen payoffs as in Table 1.

Table 1: Payoffs in the Credibility Game

	I is accepted by CC				I is rejected by CC			
	GT	BT	GF	BF	GT	BT	GF	BF
CC	2	2	-2	-2	0	0	0	0
CP-H	4	5	1	2	-1	0	-3	-2
CP-L	1	2	4	5	-1	0	-3	-2

Signaling Credibility Game

3.10   In the signaling game CC does not observe CP's type i.e. does not observe whether his partner producer is honest or dishonest. Instead CC receives a signal that is a function of the properties of the produced information (*TF* and *L*).

3.11   The signal is a random value from a normal distribution with parameters:

$$Mean = w_{TF} * TF + w_L * L$$

$$Standard\ deviation = \delta$$

where $w_{TF} = w_L = 1$ are weights of the *TF* and *L* properties of the produced information, and δ is a constant that models the

expertise of a consumer.

3.12 The larger the δ is, the more noise in the signal is received. In this paper, we only use two values of δ: 0.05 and 0.66. We call a consumer an *expert consumer* if he has $\delta = 0.05$. All other consumers are called *non-expert consumers*, because they have $\delta = 0.66$.

3.13 The weights w_{TF} and w_L reflect how heavily the signal is influenced by the actual reliability of information and how it is being distorted by the look. They have an important role in our model. They can be used in modeling improvements of the credibility judgments of information consumers by increasing the role of TF in the signal. A perfect credibility judgment would be achieved if the signal depended only on TF. At the other extreme, naive consumers take into account only the presentation quality L of the Web page, instead of its truthfulness TF. In between these two extremes, several versions of the signal can be created that depend on both TF and L. In our simulation we use the following weights: w_{TF}=0.33 and w_L=0.66 to model situation where users prefer good-looking information regardless its truthfulness, and w_{TF}=0.66 and w_L=0.33 to model the opposite situation. Admittedly this covers only a fraction of all possibilities. It also seems natural to vary the weights across consumers of different expertise. We leave these extensions for future work.

3.14 *The strategy of CC in our basic model is based on a threshold for the signal value*: if the signal value is above the threshold, CC accepts the produced information, if it is below the threshold, CC rejects the information.

Credibility Game With Reputation

3.15 We decided to enhance our basic Credibility Game with a reputation system. The system we use is similar to the ones utilized by many existing web services and news aggregators. Consumers can leave feedback by voting up or down, which corresponds to adding or removing one point of reputation from the CP's score respectively. Total reputation score is a plain sum of the votes. The role of this score is twofold. Firstly, it has been incorporated into the consumer's acceptance strategy. We have investigated four possibilities: strategies, based solely on signal, solely on reputation, as well as the logical alternative and conjunction of the two. Therefore, in this extension, CC's strategy consists of a pair of thresholds: one for the signal and one for the reputation. The second role of the reputation score is related to the information visibility. We have examined two matching rules. The first one is a 100% random choice. The second approach starts with a random selection of three producers from the entire population. Among these three producers, one with the highest reputation score is then chosen.

3.16 In our simulations we investigate two voting rules that determine how consumers evaluate producers. The first one is based on signal. It models a situation where a consumer leaves feedback immediately after the initial evaluation, relying on his first impression. In this case, CC votes the producer up whenever the information signal exceeds the threshold and votes down whenever it does not. The other voting rule we chose to examine is based on game payoffs. This one illustrates the behavior of consumers, who after choosing to accept or reject the information are able to verify whether relying on the information has been beneficial or not, and can leave feedback dependent on this verification. For example, if the produced Web content contains the description of a treatment for a disease, a signal-based rating is produced when a consumer finds the content to be credible without trying the treatment, while a payoff-based rating is produced if the consumer tries the treatment and reports the results to the reputation system. In the case of simulation the consumer rates, the producer only if the interaction alters his payoff. If the payoff increases, the feedback is positive. If the payoff decreases, the feedback is negative.

3.17 Details of the reputation system implementation in our simulator, along with the possible extensions to support more sophisticated mechanisms, are described in the Simulator design. The brief overviews of the rating strategies' and producer choice strategies' internal design are provided in the Rating strategy section and Matching rule section respectively.

🌐 Simulator design

4.1 The described model has been implemented using Repast Simphony 2.0 framework. In this section we provide an overview of the simulator design and mention some of its possible applications for future research. Additionally, in Table 3 we present a list of the most important parameters, which can be used to model different environment settings and simulation scenarios. Appendix A contains a detailed description of the simulator design.

Overview

4.2 Each simulation run involves several basic steps, namely initializing the populations of producers and consumers, performing certain number of interactions between each consumer and the chosen producer, evaluating payoffs and changing the properties of both types of agents. The detailed procedure is described below, along with some important parameters that need to be set. The parameters are briefly described in the Parameter settings for reported simulations section.

1. Generate initial population of producers *[producer_population_size]* and consumers *[consumer_population_size]*.
2. For each generation *[generation_number]* do
   a. For each iteration *[iteration_number]* do
      i. Each producer

1. produces one piece of information according to his strategy
ii. Each consumer
1. chooses one producer from the population according to the producer choice strategy
2. evaluates the producer's information and decides whether to accept it or not
3. modifies its gain depending on the accepted/rejected information features and payoffs *[consumer_payoffs]*
4. rates the producer according to his rating strategy
(this action only modifies the working copy of the producer's rating, so that other consumers do not take this evaluation until next iteration)
iii. Gains are increased/decreased for all the producers depending on their type (honest/dishonest), information features (truthfulness and look) and its acceptance rate, payoffs *[producer_honest_payoffs]* *[producer_liar_payoffs]*
iv. The original producers ratings are updated taking into account the evaluations from this iteration
b. Producers evolve (modify the TF and L properties of the information produced in the next generation)
c. The initial ratings of all producers are reset*
d. Consumers evolve (modify their acceptance thresholds for both signal and reputation)

*Note that the reputation is valid only for all iterations of the producer->consumer interactions within one generation, and affects the survival of particular strategies.

Producer Strategy

4.3   In the simulations reported in this paper, a producer's strategy was limited to choosing among the available actions BF, GF, BT, GT. However, the simulator allows us to define and study more complex strategies, for example making the choice of an action dependent on current reputation level.

Consumer Strategies

Matching rule

4.4   The existing simulator implements two possible strategies regarding the choice of the producers for interaction from the entire population (see Producer choice strategies class diagram). The first one allows us to choose one producer at random. The second one randomly picks three candidates and then chooses the one with the highest reputation. However this predefined set can easily be extended to model more sophisticated strategies.

Acceptance strategy

4.5   Currently there are two factors that can be taken into account by consumers, while deciding whether to accept information or not. These are acceptance by signal (using the signal threshold) and acceptance by reputation (using the reputation threshold). One may choose to use only one of them; however, it is also possible to incorporate logical alternative or conjunction of both. The type of strategy is predefined for all the consumers at the beginning of the simulation by setting the *[consumer_acceptance_strategy]* parameter. The parameter values along with the corresponding acceptance strategies are shown in Table 2.

Table 2: Acceptance strategies and the corresponding parameter values

Strategy	Parameter value
By signal only	0
By reputation only	1
By signal and reputation	2
By signal or reputation	3

Evolution of Agent Strategies

4.6   Evolution of the producers' strategies involves altering the truthfulness and look of the generated content. We apply the Stochastic Universal Sampling technique to select the strategies potentially useful for recombination, using payoffs as a fitness metric. This method allows us to avoid genetic drift (Samuelson 1997). It is important to note, that the producers can only copy strategies from other producers within the same type (namely honest or dishonest). The detailed procedure consists of the following steps:

1. Divide the population into groups of honest and dishonest producers

2. For each group
    a. Use Stochastic Universal Sampling to select strategies that are going to be copied
    b. For each producer in the group
        i.Copy the truthfulness and look value from the selected strategy
3. Alter the look strategy of 1% of randomly chosen consumers to a random value (0 or 1)

4.7   For consumers, only the acceptance strategy can evolve. This involves altering signal and reputation thresholds by copying strategies of the best consumers and adding small mutation to it. We apply the Stochastic Universal Sampling technique to select the strategies potentially useful for recombination, using payoffs as a fitness metric. It is important to notice, that consumers can only copy strategies of other consumers within the same type (namely with the same expertise level). The detailed procedure has similar steps as for the evolution of producer strategies.

Parameter settings for reported simulations

4.8   In the table below we describe the most important simulation settings for all the results presented in the section Impact of Reputation on Credibility Evaluation.

Table 3: The most important simulation settings

Parameter	Values	Additional info
generation_number	100	These settings are fixed for all
iteration_number	100	the scenarios described in the
consumer_population_size	1000	Impact of Reputation on Credibility Evaluation
producer_population_size	100	section
signal_truthfulness_weight	0.33	
signal_look_weight	0.66	
expertise_levels*	0.05, 0.66, 0.66, 0.66, 0.66, 0.66, 0.66, 0.66, 0.66, 0.66	
consumer_payoffs	0,0,0,0,-2,-2,2,2	
producer_honest_payoffs	-2,-3,0,-1,2,1,5,4	
producer_liar_payoffs	-2,-3,0,-1,5,4,2,1	
producer_liar_rate	0.8	
	0.2	
producer_reset_threshold	0	Environment with 80% dishonest producers (see section Environment with mostly dishonest producers)
	80	Environment with 80% honest producers (see section Environment with mostly honest producers)
producer_choice_strategy	CredibilityGame.producerchoice. strategy.RandomProducerStrategy	Scenarios without reputation whitewashing (see section No whitewashing)
	CredibilityGame.producerchoice. strategy.ByReputationProducerStrategy	Scenarios involving reputation whitewashing (see section 80% whitewashing)
consumer_rating_strategy	CredibilityGame.rating.strategy. SignalDependentRatingStrategy	Choosing one producer from the entire population at each simulation step (referred to as one random in the tables with results)
	CredibilityGame.rating.strategy. PayoffDependentRatingStrategy	Choosing producer with the highest reputation from three randomly chosen ones at each simulation step (referred to as best reputation from three random in the tables with results)
consumer_acceptance_strategy	-1	Environment, where consumers leave positive feedback, if information has been accepted by signal, and negative otherwise (see section Signal-based reputation system)
	0	Environment, where consumers leave positive feedback, if they have increased their gains, and negative if their gain was decreased (see Payoff-based reputation system section)
	1	At the beginning of the simulation, acceptance strategy is chosen at random from all the available

	ones - it can be altered during the evolution process (see section Strategy types confrontation)
2	Acceptance based solely on the signal threshold (referred to as *by signal* in the tables with results)
3	Acceptance based solely on the reputation threshold (referred to as *by reputation* in the tables with results)
	Acceptance based on the logical conjunction of acceptance by reputation and by signal (referred to as *by reputation and signal* in the tables with results)
	Acceptance based on the logical alternative of acceptance by reputation and by signal (referred to as *by reputation or signal* in the tables with results)

* Expertise level of each consumer is chosen at random from this predefined table with equal probability. It is used while computing the signal (see section Signal for details) – in general the lower value yields more accurate signal. In the examined setting we only had two expertise levels, however the higher value was chosen with higher probability. Different settings, regarding various combinations of consumers' expertise and signal properties, have been considered in Abramczuk et al. (2012).

Impact of Consumer Expertise on Credibility Evaluation

5.1 In Abramczuk et al. (2012) we have experimentally analyzed the evolution of agent strategies in the signaling Credibility Game under producer population mixes of varying honesty and under consumer population mixes of varying expertise. Moreover, we have varied the transparency of the truthfulness signal to the consumers to assess cases where various presentation tricks can be employed by dishonest producers to mislead consumers towards information acceptance.

5.2 Simulation results for the signaling game with TF strategy evolution indicate that the signal itself is not sufficient to alter the dishonest producers' behavior and provide incentives to submit true information in an environment with only less experienced consumers, who exhibit preference for look (look weight=0.66). Figure 1 shows the exact trace of the dishonest producers' strategies changes, against different populations of consumers and signal properties in the environment, where most of the producers (80%) are dishonest. The presented values are averages of strategy proportions in last generation from 10 simulation runs.
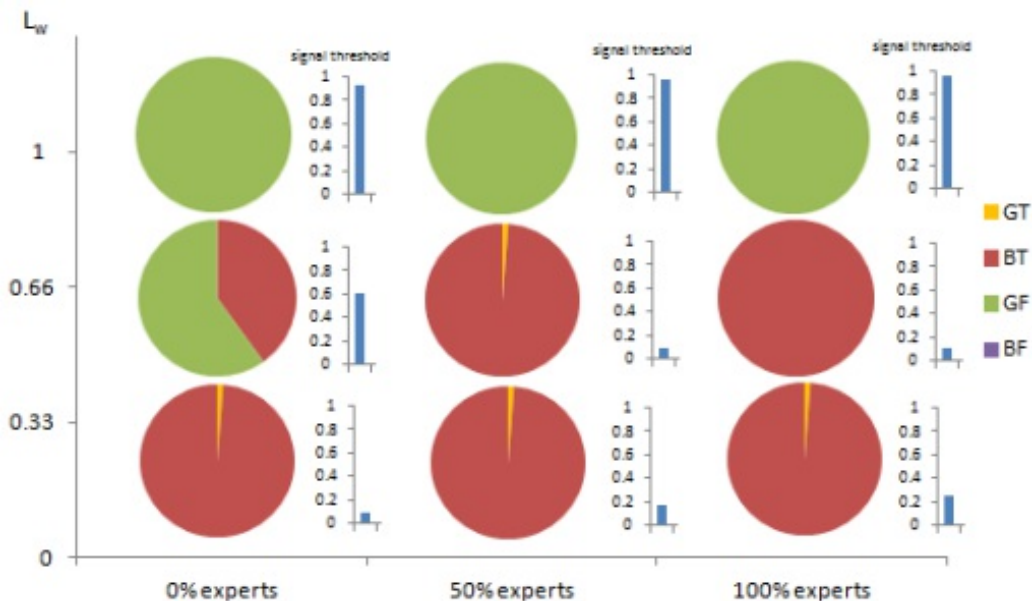

Figure 1. Evolution of dishonest producers strategies

5.3 The effects shown by our simulation study are not always stable[2], as in the less extreme simulation settings some simulation runs exhibit varying behavior. However, *the results clearly support hypothesis 1 in the case when there are no experts* (all consumers receive a signal that is based mostly on look).

5.4 We have evaluated this hypothesis by performing two-sample Wilcoxon rank-sum (Mann-Whitney) test, investigating the statistical significance of the difference between the behavior of dishonest producers in an environment with 0% expert consumers and 50% expert consumers (signal look weight=0.66). The tested hypotheses were defined as follows:

h_0: The same number of dishonest CP choose to produce true information in an environment with 0% expert consumers and in an environment with 50% expert consumers (difference=0).

h_a: More dishonest CP choose to produce true information in an environment with 50% expert consumers, than in an environment with 0% expert consumers (difference>0)

5.5   Assuming the significance level equal to 0.05, there is sufficient evidence to reject the null hypothesis (p= 0.0043).

🌐 Impact of Reputation on Credibility Evaluation

6.1   In the previous section, we have shown that when consumers lack expertise, dishonest producers tend to exploit them and to provide false, but well-presented information. This phenomenon might actually be a problem in a real-life system, as it is very likely that a large fraction of users, who seek information on the Internet, are not the most experienced ones (Hargittai 2002), and therefore can only evaluate content basing on the first impression. However, a great variety of the existing web services allow the community members not only to post information, which is the activity of content producers in our model, but also to leave feedback for other users. The latter is a crucial extension of the consumer's strategy, which can have a significant influence on the overall system performance.

6.2   The Credibility Game has been extended with reputation to investigate the impact of reputation on CP's behavior. We have used this extension to examine several hypotheses regarding the influence of reputation on the community, where the fraction of expert consumers is extremely low (namely 10%) and information look is very important for all consumers. Such scenario can be plausible for many publicly available services, as it might be difficult to provide experts with incentives to actively participate in such a community. Our main goal was to investigate, whether different acceptance and voting rules can effectively improve information truthfulness against populations of mostly honest or mostly dishonest producers. Additionally, we aimed to examine the influence of reputation whitewashing on the results.

Signal-based reputation system

Environment with mostly honest producers

6.3   We have compared the results with data generated in the setting, which did not involve reputation at all (also referred to as reference results). In this scenario, consumers accepted information solely based on signal and chose the producers for interaction at random. For such a setting we have observed, that all of the honest producers not only decided to generate true information, but also to invest in presentation. On the other hand, all of the dishonest ones invested in presentation; however, had no incentive to provide truthful information. These results were 100% compatible with the signal properties.

No whitewashing

6.4   We have observed that, regardless the acceptance rule, reputation provides an incentive for dishonest producers to switch to the GT strategy (which according to their payoffs is less profitable in case of acceptance, but safer in case of rejection), if reputation is taken into account, while choosing the producer for interaction (see Figure 2). At a first glance, it may seem that this phenomenon is caused by the environment of mostly non-experts where a relatively large standard deviation gives significantly bigger chances of increasing reputation to the GT strategy. Therefore, worse strategies are not chosen for interaction with the consumers frequently enough to be profitable, even if the consumers are not experienced and there is a good chance of accepting false information.
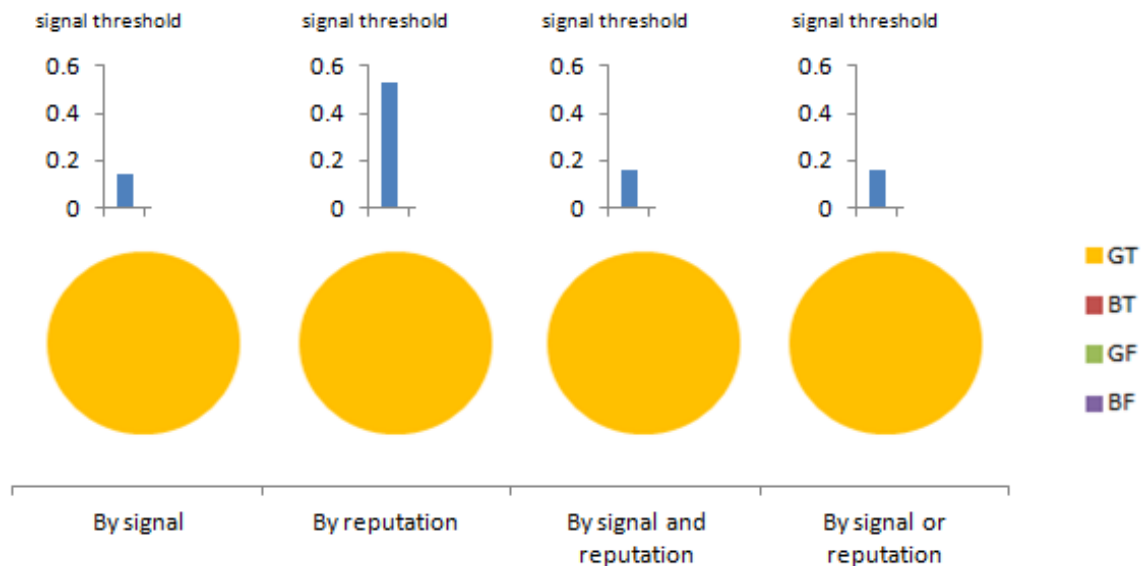
Figure 2. Results for reputation-based producer choice

6.5   However, the results suggest a different explanation. The average number of consumers, who were given the GF information for evaluation in the first generation of the reference results, is about 26% of the entire population, while after changing the producer choice strategy to the reputation based one, this number increases to about 39%, which is probably caused by filtering out some of the weakest (in terms of signal properties) strategies, namely BF and BT. This suggests, that even with a relatively small number of experts in the community and the emphasis on presentation in the signal properties, dishonest producers do have incentive to produce true information, if they are exposed to the consumers' evaluation frequently enough. Moreover, only four generations were required to entirely eliminate the GF strategy from the population of producers (both honest and liars).

6.6   The generalized results of the final simulation outcome are as follows:

6.7   When reputation is taken into consideration, while choosing the producer for interaction, all the producers choose GT regardless the acceptance strategy. When the producers are chosen at random, the results depend on the acceptance strategy:

- Acceptance by signal only (reference results), acceptance by reputation or signal
    °GT is chosen by all the honest producers
    °GF is chosen by all the dishonest producers
- Acceptance by reputation only, acceptance by reputation and signal
    °GT is chosen by all the producers

Table 4: Consumer acceptance strategies in last generation

Acceptance strategy	By signal AND reputation	By signal OR reputation
Average expert's signal threshold	0.35	0.77
Average total signal threshold	0.15	0.21
Average reputation threshold	324.93	1619.4
Average reputation	810	714.33

6.8   We can observe the same results (namely choosing the GT) for the random producer choice, if the acceptance strategy is based solely on reputation or on reputation *and* signal. On the other hand, the results for a random producer choice and the acceptance strategy based on signal *or* reputation are similar to the reference results (where the acceptance strategy is based solely on signal and the choice of a producer for interaction is 100% random). In this case, it is not profitable for the dishonest producers to generate true information, as they can exploit signal properties to increase their acceptance rate and, according to payoffs, the profits from a single interaction with one consumer are (in case of acceptance) four times larger for the GF strategy than for the GT strategy.

6.9   We have performed a more thorough analysis of the results involving two acceptance strategies. The first one was based on signal *and* reputation, while the other one relied on signal *or* reputation. The investigation of the consumers' strategies in the last generations is shown in the table *above*.

6.10   The acceptance strategy based on signal *or* reputation seems to be less restrictive, however the final results of the simulation reveal, that it results in a slightly higher overall signal threshold (while significantly higher for the minority of experts) and an extremely high reputation threshold. On the contrary consumers, whose acceptance strategy is based on the theoretically more

restrictive logical *and* operator end up with lower signal and reputation thresholds. This phenomenon is probably caused by frequent errors in the early-stage generations and the consumers' payoffs setting. The logical and operator applied to the acceptance by signal *and* by reputation results in only 25% chance of accepting the processed information (there are four possible configurations and only one of them yields the overall acceptance), while at the beginning of the simulation 50% of strategies generate true information. This, along with the low expertise of the consumers, probably results in the high rejection rate of true information, which does not allow the consumers to increase their payoffs. Thus, the ones with the lower thresholds are preferred in the evolutionary process, as they have a better chance to increase their payoffs. On the contrary, the logical *or* operator nurtures high acceptance rate at the beginning of simulation, and therefore allows the consumers to learn by "punishing" the acceptance of false information. This results in a dramatic increase of both thresholds values (reputation threshold in particular). Interestingly, in this scenario the final results indicate that the inflated reputation threshold did not provide dishonest producers with an incentive to generate true information. By the end of the simulation, dishonest producers all chose the GF strategy, while in the previous setting with the lower final reputation threshold they decided to alter their behavior and produce accurate content. This suggests that when the reputation threshold of the consumers passes some critical value, it becomes extremely difficult for the producers to get accepted by reputation, and therefore, they choose to comply with the signal properties. Moreover, the GF strategy is, according to payoffs, more beneficial for them in case of acceptance.

80% whitewashing

6.11   We decided to check whether the results presented above will change if we allow reputation whitewashing. If not discouraged, it can become a fairly common practice for dishonest producers to abandon their discredited identities and take on new ones. This behavior can potentially lead to substantial lowering of reputation system efficiency.

6.12   To model the phenomenon of reputation whitewashing, we have allowed the producers to randomly reset their reputation at the end of each iteration. The probability is specified by an additional parameter (it is 80% for the described scenarios with whitewashing). We have examined how this phenomenon affects simulation outcome.

6.13   When the acceptance strategy was based only on signal and the matching rule was based on reputation, the results were exactly the same as the reference results. This is surely caused by the producer choice randomization which constitutes a direct effect of 80% whitewashing.

6.14   The results for the acceptance strategies based on reputation and signal, and on reputation only, seem to be a little bit surprising. Regardless the matching rule, 80% whitewashing did not remove the incentive to produce true information (in case of random producer choice it even provided the motivation to do it). However, it did remove the incentive to invest in presentation (for both producer types). In-depth analysis of the final population reveals that the dominant strategy chosen by almost all producers of both types has a 0% acceptance rate, as an average rating of about 4 at maximum never gets close to the average threshold. This low acceptance rate explains the strategy choice because, according to payoffs, only producing BT information allows the producers not to lose anything.

6.15   The results for the by signal *or* reputation type of strategy are the same as the reference results. This is not particularly surprising when compared with the same setting without whitewashing. In that case, reputation was practically irrelevant for the dishonest producers evolution, because it was more beneficial for them to remain compliant with the signal properties. Therefore, even 80% chance of whitewashing has no influence on the final population.

Environment with mostly dishonest producers

6.16   Next we decided to analyze the reputation system behavior in populations in which dishonest producers dominate. We have compared the results with data generated in the setting, which did not involve reputation at all (also referred to as the reference results). In this scenario consumers accepted information solely based on signal and chose the producers for interaction at random. For such setting, we observed, that the results are not 100% consistent. Eight out of ten simulation runs finished with all the producers choosing to provide true information. However, only the honest ones deciding to invest in presentation (BT strategy is the best choice in case of frequent rejection). In some cases though, the dishonest producers still preferred the GF strategy. These differences were not present in the environment with mostly honest producers where it was less likely for an expert consumer to interact with a dishonest producer. The results indicate that solely increasing the chance of interaction with the consumers may in some cases provide the dishonest producers with an incentive to generate true information.

No whitewashing

6.17   It seems that regardless the acceptance strategy, reputation provides the incentive for the dishonest producers to switch to the least profitable (in the case of acceptance) strategy of producing GT information.

6.18   When the acceptance strategy is based only on reputation and the matching rule is random, the results are the same as described above (all or almost all producers choose to produce GT information.

6.19   For all the other acceptance strategies (involving only signal and both mixes of signal and reputation) the results of random

producer choice are slightly inconsistent.

6.20   The acceptance strategy by signal and reputation results in the dominance of GT information in nine simulation runs out of ten. However, in one case this strategy is chosen almost only by honest producers. Still, even in this situation, the dishonest producers choose to produce true information but do not invest in presentation, which suggests a high rejection rate for this type of producers.

6.21   For the acceptance strategy by signal or reputation the final results are 100% consistent for honest producers who, as usual, tend to choose the GT strategy. The dishonest ones choose to produce BT in most cases (seven out of ten runs). Nevertheless, three simulation runs finish with the dominance of GF information among them. This indicates that in most cases reputation dominates the signal and improves the truthfulness in the system, although this is not a general rule. The average ratios of each strategy in the last generation for this scenario are shown in Figure 3.
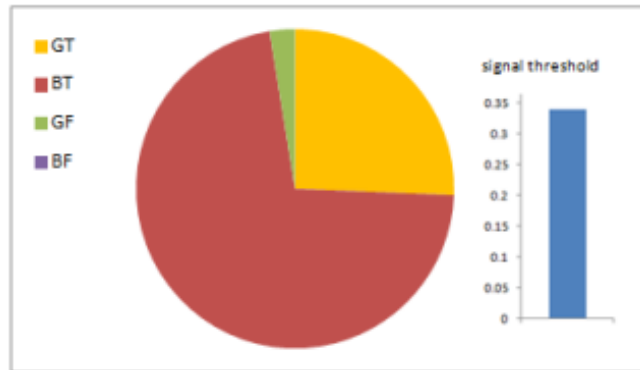

Figure 3. Ratings based on signal, random producer choice, acceptance by reputation or signal

80% whitewashing

6.22   Further, we investigate the influence of whitewashing. When the producer choice strategy is reputation-based and the acceptance strategy is based on signal only or on signal or reputation, the results are inconsistent.

6.23   However, when the acceptance strategy is based solely on reputation or both on reputation and signal, 80% whitewashing does not have a negative impact on the truthfulness of information produced by both types of producers, regardless the producer choice strategy. It does though remove the incentive to invest in presentation. The reasons are the same, as for the same scenario in an environment with mostly honest producers – 80% whitewashing practically makes it impossible for the information to get accepted, due to the extremely low ratings in comparison with the reputation threshold. Thus all the producers choose the strategy, which does not involve additional costs in terms of payoffs.

Payoff-based reputation system

6.24   In the results presented above we were assuming that producer ratings are based on consumers' evaluations of reliability of information that the producers offer. In other words, the CC voted the CP up whenever he accepted the information and voted CP down whenever he rejected the information. In case of reputation based acceptance strategies this may lead to simple herding. However, it can be argued that consumers should rate producers only when they have first-hand experience with the reliability of a given information piece i.e. they accepted it and know how it changed their payoffs. In this section we present analysis of this situation.

Environment with mostly honest producers

6.25   Similarly to the signal-based reputation system, we have compared the results with data generated in the setting, which did not involve reputation at all (also referred to as reference results). In this scenario, consumers accepted information solely based on signal and chose the producers for interaction at random. In such setting we have observed that all of the honest producers not only decided to generate true information, but also to invest in presentation. On the other hand, all of the dishonest ones invested in presentation, although they had no incentive to provide truthful information. These results were 100% compatible with the signal properties.

6.26   Similarly to the signal-based reputation system, we have compared the results with data generated in the setting, which did not involve reputation at all (also referred to as reference results). In this scenario, consumers accepted information solely based on signal and chose the producers for interaction at random.

No whitewashing

6.27   In general, the results indicate a stronger preference for the BT strategy regardless the producer type (when choice of the producers for interaction is 100% random) in comparison with the signal-based rating system. We can also observe a stronger preference for generating true information, which seems to be a natural result of a more accurate feedback based on the actual interaction results, and not on the first impression (see Figure 4).
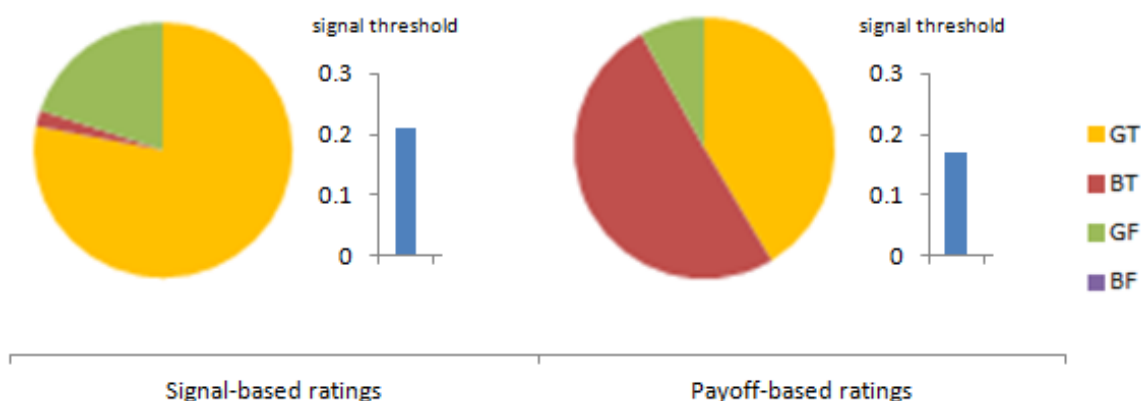

Figure 4. Acceptance by signal or reputation, random producer choice

6.28   Interestingly, producers seem to exhibit tendency not to comply with the signal properties, and more often choose not to invest in information presentation. For some acceptance strategies, this behavior can only be observed among dishonest producers. However, for instance, when acceptance is based solely on reputation, strategy <true, bad-looking> is chosen almost by the entire population of producers (see Figure 5).
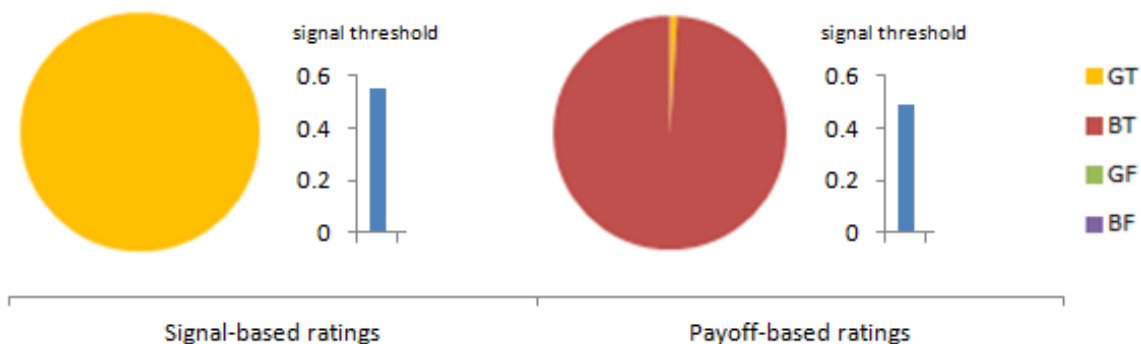

Figure 5. Acceptance by reputation, random choice of producers

6.29   Detailed results are presented in Detailed comparison of the results for ratings by signal and by payoffs (note, that all the presented values are averages from ten simulation runs with the same parameters).

80% whitewashing

6.30   When we introduce 80% probability of reputation whitewashing, the final results are almost exactly the same as the corresponding results for the reputation system based on signal.

Environment with mostly dishonest producers

6.31   Similarly to the signal-based reputation system, we have compared the results with data generated in the setting, which did not involve reputation at all (also referred to as the reference results). In this scenario, consumers accepted information solely based on signal and chose the producers for interaction at random. For such setting, we observed that the results are not 100% consistent. Eight out of ten simulation runs finished with all the producers choosing to provide true information; however, only the honest ones deciding to invest in presentation (BT strategy is the best choice in case of frequent rejection). In some cases though, the dishonest producers still preferred the GF strategy. These differences were not present in the environment with mostly honest producers, where it was less likely for an expert consumer to interact with a dishonest producer. The results indicate that solely increasing the chance of interaction with the consumers may in some cases provide the dishonest producers with an incentive to generate true information.

No whitewashing

6.32   When a reputation system is included, we get results analogous to the ones described for an environment with mostly honest producers. We can observe differences only for the scenario with acceptance strategy based on the logical alternative of reputation and signal, and random matching rule. The results are inconsistent, as they were in an environment with mostly honest producers; however, in the end, the producers are more evenly distributed, while choosing among the three final strategies, namely GT, BT and GF (see Comparison of the results for mostly honest and mostly dishonest producers).

80% whitewashing

6.33   When 80% probability of whitewashing is present, the results are very similar to the ones generated for the corresponding setting in the rating system based on signal.

Strategy types confrontation

6.34   The experiments described above, have been performed using fixed, predefined types of acceptance strategies. To verify the second hypothesis, we allowed the types of acceptance strategies to evolve along with the signal and reputation thresholds. The initial types of strategies were assigned at random with about 25% chance of assigning each one of the four, currently available in our model. We have examined the results of the strategies evolution for both signal- and payoff-based ratings, in an environment with only 10% of expert consumers and 80% of honest producers. We have observed that when the ratings were based on the first impression (signal-based ratings), the evolving acceptance strategy did not provide the producers with incentives to produce true information, while when the rating process was performed after the evaluation of payoffs (payoff-based ratings), all the dishonest and almost half of the honest producers chose to produce BT information (see Results for the evolving acceptance strategies).

6.35   In both cases, the majority of consumers chose an acceptance strategy based on the logical alternative of reputation and signal, and the rest of the consumers accepted information based solely on signal. Preference for the first strategy is significantly larger for the payoff-based ratings (see Table 5).

Table 5: Acceptance strategies chosen by the consumer

	Signal-based ratings	Payoff-based ratings
By signal only	*~20% consumers*	*~5% consumers*
By signal or reputation	*~80% consumers*	*~95% consumers*

6.36   *These results confirm hypothesis 2, which stated that when reputation is preserved (no whitewashing), consumers use reputation to complement signal as a way to evaluate credibility of information*. When reputation is based on payoffs, this applies to 95% of all consumers; while when reputation is based on signal, to about 80%. We have evaluated this hypothesis using binomial test. The verified hypotheses were defined as follows:

h_0: In 80% of simulation runs at least 60% of consumers (namely 600) will incorporate reputation into their acceptance strategy.

h_a: In more than 80% of simulation runs at least 60% of consumers (namely 600) will incorporate reputation into their acceptance strategy.

6.37   Assuming the significance level equal to 0.05, there is sufficient evidence to reject the null hypothesis (p= 0. 01153).

6.38   We have also performed a detailed analysis of the evolution of both signal and reputation thresholds evolution (see Table 6). The results reveal that the reputation threshold is significantly lower, when the ratings are based on payoffs, in comparison with the signal-based ones. Inflated reputation threshold probably provides the producers with incentives to comply with the signal properties; as such behavior makes it more probable to get accepted. The general signal threshold (joint for both experts and non-experts) is similar for both rating strategies; however, in the case of signal-based ratings, an average expert threshold is significantly higher. When the signal properties assign larger weight to the information presentation (TF=0.33, L=0.66) an inflated signal threshold makes it extremely difficult for the expert consumers to make use of their knowledge and accept the BT strategy. Lower signal threshold of the non-experienced consumers does not help, because there is a high probability for them to make an inaccurate initial evaluation.

Table 6: Signal and reputation threshold evolution

Strategy	Signal-based ratings	Payoff-based ratings

Average reputation threshold	1632	1073
Average rating	726	805
Average signal threshold (overall)	0.19	0.1
Average expert's signal threshold	0.76	0.1

6.39 Apart from the previous analysis, we have also investigated, what is the influence of particular rating strategies on the overall system performance (namely the average total payoffs of consumers and producers) when we allow the consumer acceptance strategies to evolve. The results shown in Overall system outcome in terms of payoffs indicate that when the acceptance is based solely on reputation, payoff-based ratings do have positive influence on the general outcome. They slightly increase payoffs of consumers and yield significantly better outcome for the producers, as they are not forced to invest in presentation.

6.40 *These results confirm hypothesis three: payoff-based reputation leads to a choice of the BT strategy by the majority of producers, which leads to a higher average of payoffs*. This is caused by the fact that if consumers accept BT, producers do not incur a costly investment into presentation. We have evaluated this hypothesis by performing two-sample Wilcoxon rank-sum (Mann-Whitney) test, investigating the statistical significance of the difference between the total outcome of producers and consumers, when the ratings were based on payoffs in comparison to those based on signal. The tested hypotheses were defined as follows:

h_0: The total gain of both populations will be equal regardless the rating strategy ($\text{gain}_{\text{payoff-based}}/\text{gain}_{\text{signal-based}}=1$)

h_a: The total gain of both populations is lower, when the rating strategy is based on signal ($\text{gain}_{\text{payoff-based}}/\text{gain}_{\text{signal-based}}<1$)

6.41 Assuming the significance level equal to 0.05, there is a sufficient evidence to reject the null hypothesis (p= 0.0002).

Conclusions and Future Work

7.1 In this paper, we have presented the Credibility Game, a game-theoretic and simulation model for the study of Web content credibility or quality. The proposed model allows for an explicit modeling of various properties of information; we have focused on two such properties, truthfulness and look, which can be thought of as jointly modeling the content of information that has an impact on payoffs of information consumers, and of the presentation of the information. Our model allows us to capture an essential property of Web content evaluation: a superficial evaluation on the basis of a signal, rather than the expected payoff from the obtained information content. The signal in the Credibility Game is a function of information properties and can be used to model consumer expertise.

7.2 We have also extended the basic Credibility Game with reputation, which allows information consumers to use another, realistic and relevant mechanism for evaluating information credibility. Our reputation model is similar to simple reputation systems widely used on the Web today.

7.3 Equipped with such a simulation model, we have set out to verify the following hypotheses:

1. Consumer expertise is a critical factor that affects the choice of producers whether to produce truthful information and whether to invest in presentation.   a. Dishonest consumers take advantage of the absence of reputation and consumer expertise to produce false information.
2. When consumers receive an inaccurate signal about Web content, and reputation is preserved (no whitewashing), consumers use reputation to complement signal as a way to choose producers.
3. Reputation (in absence of whitewashing) that is based on payoffs gives better results than reputation based on signal for the whole community (higher payoffs for both producers and consumers).

7.4 We have observed, that when the truthfulness weight is extremely low, dishonest producers indeed take advantage of the lack of experts in the population and keep generating false information. Increasing the number of expert consumers may provide the dishonest producers with an incentive to alter their behavior; however, only when they are exposed to the interaction with the consumers frequently enough. Consumers rely strongly on a signal, while making the acceptance decisions. Nevertheless, when we allow the acceptance strategy to evolve, most of them choose the logical alternative of both reputation and signal (this effect is stronger, when consumer ratings are based on payoffs). Moreover, the analyzed results indicate, that payoff-based ratings yield better performance for the entire system in terms of the total average payoffs.

7.5 The proposed model allows us to make a consideration of several phenomena that play an important role in the selection of high-quality content in today's Web-based information systems. Signal-based reputation ratings, in particular, have a strong resemblance to Facebook likes, Reddit upvotes or other collaborative filtering mechanisms used in social media today. Using the Credibility Game, we have been able to study the effectiveness of signal-based reputation for web users, and to compare it against payoff-based ratings. Moreover, in our model consumer strategies and producer strategies evolve separately, allowing producers to respond to the use of reputation by consumers and to choose a strategy that does not invest into presentation, but

relies on the information content itself, in order to attract consumers.

7.6　Clearly, on the Web today, both content quality and presentation play an important role in credibility judgments. *Mechanisms that aim to support and improve evaluations of Web content quality or credibility can be studied using the Credibility Game, and their effectiveness can be evaluated using simulation.* Such an approach would, for example, enable us to study the robustness of such strategies to adversaries. Another interesting issue would be the exchange of the roles of content producer and consumer, such as in typical Web2.0 systems like Question-and-Answer forums (for example, Quora or StackOverflow). Again, our model allows for such a study, but also makes it possible to consider new effects, such as the role of producer investment into presentation in consumer learning. Apart from further simulational studies of the influence of more complex reputation systems and adversary strategies on certain communities, we also plan to perform behavioral experiments regarding the most interesting scenarios. Such issues will be the subject of our future work.

Appendices

Appendix A Details of simulator design

8.1　This appendix describes the design of our simulator in more detail. The simulator code can be found at http://www.openabm.org/model/3878/version/1/view. We begin with an introduction of the most important classes.

Information

8.2　Information represents the content in the modeled system. It is associated with the strategies of its producer and all the consumers, who evaluated it. Information has only two properties, namely truthfulness and look. In our experiments, the values of both properties are binary; however, the simulator design is capable of supporting values from a continuous scale as well. The Information object is responsible for internally computing payoff modifications for both producers and consumers, depending on their acceptance decision and predefined payoffs. Detailed list of the implemented methods is shown in Information class diagram.

Producers

8.3　There are two types of producers in the system, namely honest and dishonest. To model different incentives for participation, each group has a separate set of configurable payoffs – *[producer_honest_payoffs]* for the honest producers, and *[producer_liar_payoffs]* for the dishonest ones. The initial strategies are chosen at random to ensure about 50% share of the true information, and about 50% share of the good-looking one. It is crucial to note, that the truthfulness strategy does not have to be compatible with the producer type, to allow behavior changes during the evolution process.

Signal

8.4　Signal is a random value, which depends on the information properties (namely truthfulness and look) and the consumer's expertise level. The main properties of this value are listed in the Table 7.

Table 7: Signal properties

Property	Value	Range of values
Distribution	Normal	
Mean	*t * truthfulness + l * look* where: t =*[signal_truthfulness_weight]* l =*[signal_look_weight]*	<0, 1>
Standard deviation	*expertise level*	<0, 1>

Ratings

8.5　Each producer stores an abstract Rating object, which in the proposed model encapsulates his total score of up/down votes. It defines an aggregation operator and can easily be extended to support some more sophisticated recommendation models. Ratings object class diagram depicts some implementation details. Current simulator supports only a very simple up/down vote rating system inspired by services like Dig[3] or Reddit[4]. We chose this one, because such solution is particularly popular in the real-life content recommendation systems. In our simulations each up vote is associated with reputation increase by one point. Reputation is symmetrically decreased after receiving a negative feedback. However, the implemented structure allows us to

model different mappings of votes to reputation modifications. This feature may be crucial while simulating services similar to Stack Exchange[5] Q&A websites, where reputation system is tuned to particularly promote certain activities.

Rating Aggregation Operator

8.6   It is used to aggregate reports from different consumers about the particular producer into one final score. In our basic rating system aggregation operator simply provides a sum of up and down votes.

Setting the Initial Ratings

8.7   By default, the initial ratings are set to 0. However, they can be configured using two parameters namely *[producer_rating_l]* and *[producer_rating_h]*. The procedure of setting the initial rating consists of the following steps:

1. Generate random number from range <*producer_rating_l*, *producer_rating_h*> (using uniform distribution)
2. Set the producer's initial rating to that number

Consumers

8.8   Several types of consumers can be identified, depending on their expertise levels *[expertise_levels]*. However, unlike the producers, their payoffs *[consumer_payoffs]* are type-independent. Expertise levels are used while computing signals – the details of this procedure are described in the Signal section. It is assumed, that the lower values are assigned to experts. It is possible to preconfigure the environment, so that it contains certain number of consumers with higher and lower expertise.

Rating strategy

8.9   Currently, the consumer can rate information, taking into account several different criteria, namely payoffs, acceptance of the information by this particular consumer (overall or signal-/reputation-based) or chosen information features. The first criterion can be used to model systems where consumers do not rate immediately after the acceptance or rejection, but verify, whether the decision to accept or reject was beneficial. The second one models scenarios, where a positive or negative feedback is provided immediately after the initial acceptance decision. The latter may be utilized for instance to create simple adversary models, where some consumers tend to rate positively or negatively information provided by certain producers. Additionally, it is also possible to model different probabilities of leaving feedback by a consumer. The implementation supports three rating strategies depicted in Rating strategies class diagram. The payoff-based one assumes that the consumer rates only, if the interaction alters his payoff. If the case of the payoff increase, the feedback is positive. On the contrary, payoff decrease results in a negative rating. In the case of random rating strategy, the producer has 50% chance of receiving a positive feedback, regardless the outcome the interaction. The signal-based rating strategy allows the producer to receive a positive rating only if the information was accepted by signal, regardless the overall acceptance. The last two strategies assume that the consumers rate every interaction with the producer.
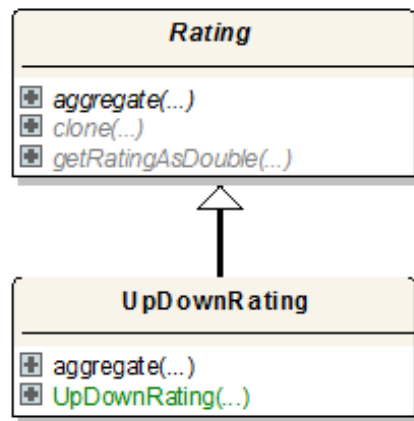
Class diagrams for the Simulator design section

Information class diagram
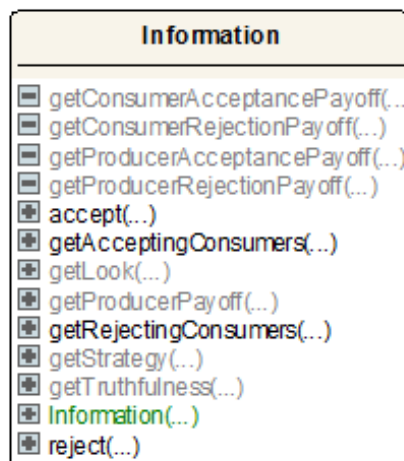
Back to the Information section.

Ratings objects class diagram

::CredibilityGame.rating

Rating
- ⊞ *aggregate(...)*
- ⊞ *clone(...)*
- ⊞ *getRatingAsDouble(...)*

UpDownRating
- ⊞ aggregate(...)
- ⊞ UpDownRating(...)

Back to the Signal section.

Producer choice strategies class diagram

::CredibilityGame.producerchoice.strategy

ProducerChoiceStrategy
- ⊞ *choose(...)*

RandomProducerStrategy
- ⊞ choose(...)

ByReputationProducerStrategy
- ⊞ choose(...)
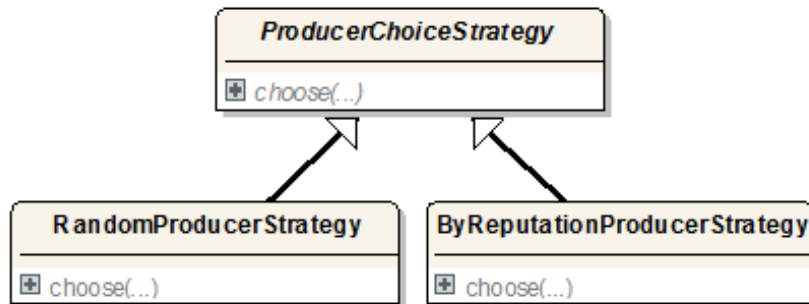
Back to the Matching rule section.

Rating strategies class diagram

::CredibilityGame.rating.strategy

RatingStrategy
- ⊞ *isForRating(...)*
- ⊞ *rate(...)*

PayoffDependentRatingStrategy
- ⊞ isForRating(...)
- ⊞ rate(...)

RandomRatingStrategy
- ⊞ isForRating(...)
- ⊞ rate(...)

SignalDependentRatingStrategy
- ⊞ isForRating(...)
- ⊞ rate(...)
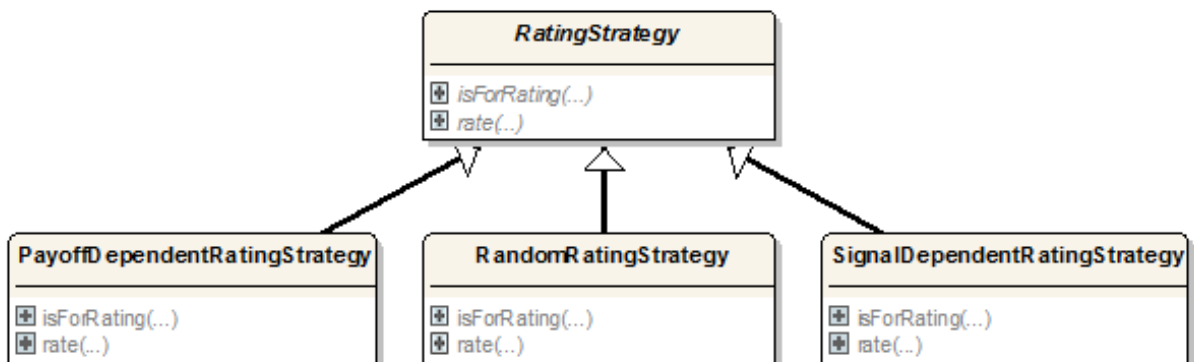
Back to the Rating strategy section.

Appendix B Details of simulation results

Detailed comparison of the results for ratings by signal and by payoffs

		Ratings based on signal				Ratings based on payoffs			

Strategy *Acceptance* By signal only

Producer choice

		One random (reference results)		Best reputation from three random		One random (reference results)		Best reputation from three random	
Truth	Look	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate
1	1	78	100%	100	80%	79	100%	100	100%
1	0	2	100%	0	0%	1	90%	0	10%
0	1	20	0%	0	0%	20	0%	0	0
0	0	0	0%	0	0%	0	0%	0	0

Strategy *Acceptance* By reputation only

Producer choice

		One random		Best reputation from three random		One random		Best reputation from three random	
Truth	Look	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate
1	1	100	80%	99	80%	1	100%	1	100%
1	0	0	0%	1	0%	99	80%	99	80%
0	1	0	0%	0	0%	0	0%	0	0%
0	0	0	0%	0	0%	0	0%	0	0%

Strategy *Acceptance* By signal and reputation

Producer choice

		One random		Best reputation from three random		One random		Best reputation from three random	
Truth	Look	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate
1	1	99	80%	100	80%	79	100%	99	80%
1	0	1	0%	0	0%	21	4%	1	11%
0	1	0	0%	0	0%	0	0%	0	0%
0	0	0	0%	0	0%	0	0%	0	0%

Strategy *Acceptance* By signal or reputation

Producer choice

		One random		Best reputation from three random		One random		Best reputation from three random	
Truth	Look	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate	*Producers*	Honest rate
1	1	78	100%	100	80%	41	100%	100	80%
1	0	2	100%	0	0%	50	76%	0	0%
0	1	20	0%	0	0%	8	0%	0	0%
0	0	0	0%	0	0%	0	0%	0	0%

Back to the results overview .

Results for the evolving acceptance strategies

Strategy		Signal-based ratings		Payoff-based ratings	
Truth	Look	Producers	Honest (%)	Producers	Honest (%)
1	1	79	100	47	100
1	0	1	100	53	62
0	1	20	0	0	0
0	0	0	0	0	0

Back to the results overview.

Overall system outcome in terms of payoffs

Strategy		*Acceptance*	By reputation only, one random			
			Signal-based rating		Payoff-based rating	
Truth	Look		*Producers*	*Honest rate*	*Producers*	*Honest rate*
1	1		100	80%	1	100%
1	0		0	0%	99	80%
0	1		0	0%	0	0%
0	0		0	0%	0	0%
Consumers' gain (average)			*131*		*133*	
Producers' gain (average)			*1900*		*2886*	

Back to the results overview.

Comparison of the results for mostly honest and mostly dishonest producers

Strategy		*Accept/reject*	Mostly honest producers		Mostly dishonest producers	
		Producer choice	By signal or reputation			
			One random			
Truth	Look		*Producers*	*Honest rate (%)*	*Producers*	*Honest rate (%)*
1	1		41	100%	11	93%
1	0		50	76%	27	11%
0	1		8	0%	16	0%
0	0		0	0%	0	0%

Back to the results overview.

Appendix C Analysis of Payoffs of the Credibility Game

9.1   The payoffs of the Credibility Game can be analyzed by considering a new game, which is constructed from the Credibility Game by an assumption that the player plays both roles of the Credibility Game simultaneously: she is a Content Producer (CP) and a Content Consumer (CC) at the same time.

9.2   Notice that a player in such a game can play any combination of the actions available for the CC and the CP, namely: BF+R, BF+A, GF+R, GF+A, BT+R, BT+A, GT+R, GT+A. For the sake of this analysis, let us focus on a combination of the most non-cooperative and most cooperative behaviors of both players. These are the combinations: BF+R (most non-cooperative) and GT+A (most cooperative). For simplicity, let us call these actions: D=BF+R, C=GT+A.

9.3   We can construct the payoff table of such a simultaneous, combined Credibility Game by adding the payoffs of the CC and the CP (keeping in mind that the player plays both roles at the same time, so he receives one payoff as a CC and one as a CP). We need to fix the CP type. Let us begin by assuming that it is CP-L (a dishonest producer).

9.4   The payoff table looks as follows:

Payoffs of simultaneous combined Credibility Game played by CC and CP-L

CC+CP-L	C=GT+A	D=BF+R
C=GT+A	3,3	-3,5
D=BF+R	5,-3	-2,-2

9.5   The game is now symmetric. Let us denote the payoffs for (C,C) by R: R=3. Also, let us denote the payoff for (D,D) by P: P=-2. Finally, let us denote the payoff of the first player for (C,D) by S: S=-3, and the payoff of the second player for (C,D) by T: T=5.

9.6   Notice that this game fulfills the assumptions of the generalized Prisoner's Dilemma:

$$T=5 > R=3 > P=-2 > S=-3$$

and

$$(T+S) / 2 < R$$

This means that the simultaneous, combined Credibility Game played by CC and CP-L is a Prisoner's Dilemma. The same does not apply, if we combine a Credibility Game played by CC and CP-H. Then, the Nash equilibrium of such a game constitutes the combination of the most cooperative actions: GT+A.

Sensitivity analysis of stable strategies to game payoffs

9.7   In order to analyze the sensitivity of stable strategies of the Credibility Game to payoffs, we have introduced an economic model of the producer payoffs that allows for their better analysis.

9.8   In this model, the utility of a content producer is given by:

$$U(I) = kG(TF) - C(TF, L)$$

where k is the number of consumers that have accepted the CP's content, G is a function that describes the gain of CP, and C is a function that describes the cost of information production. Note that G is a function of TF, not L. This utility function models the fact that the marginal cost of producing information is zero: the CP bears only a fixed cost of production that does not depend on the number of consumers.

9.9   The gain of the content producer is given by:

$$G(TF) = \alpha TF + \beta$$

$\alpha > 0$ for an honest producer (CP-H), and $\alpha < 0$ for a dishonest producer (CP-L).

9.10   The cost of the content producer is given by:

$$C(TF, L) = \gamma TF + \delta L + \varepsilon$$

where $\gamma < 0$ models that increasing the truthfulness of the produced information reduces the cost of its production (the costs is determined by cognitive effort for manufacturing a falsehood). $\delta > 0$ models that improving the look increases the cost. $\varepsilon > 0$ models a fixed cost.

9.11   Consider the following parameters for CP-L:

$$\alpha = -5, \qquad \beta = 7, \qquad \gamma = -2, \qquad \delta = 1, \qquad \varepsilon = 2$$

The resulting utility function is given by:

$$U^L = \begin{cases} -3TF - L + 5, & k = 1 \\ 2TF - L - 1, & k = 0 \end{cases}$$

This utility function produces the payoffs of CP-L for the Credibility Game.

9.12   Consider the following parameters for CP-H:

$$\alpha = 1, \qquad \beta = 4, \qquad \gamma = -2, \qquad \delta = 1, \qquad \varepsilon = 2$$

The resulting utility function is given by:

$$U^H = \begin{cases} 3TF - L + 2, & k = 1 \\ 2TF - L - 1, & k = 0 \end{cases}$$

This utility function produces the payoffs for CP-H for the Credibility Game.

9.13   Notice that the cost functions for CP-H and CP-L do not differ. For this reason, in order to study the sensitivity of strategy stability to payoffs, it is sufficient to study a variation of the coefficients α and β. We have done so in the range of $-10 < \alpha, \beta < 10$. For each combination, we have run a simulation until a stable strategy emerged for the producer depending on its type. Simulation was set up for population of 100 CP and 1000 CC. Since honest producer strategy is quite straightforward, we have decided to use 80:20 mix of types where CP-L are 80% of producer population.

9.14   This created a matrix of 400 simulation results, showing strategy distribution at the end of simulation. In almost all cases, a single dominant strategy was found for each of type of the producer (CP-H and CP-L). Using simulation results, we have drawn maps where are shown dominant strategies for the investigated range of utility function coefficients. Figure 1 depicts map of stable strategies for each type of content producer. On each map, it is possible to distinguish two intersecting lines that are boundaries of areas that promote good-looking strategy vs bad looking strategy (this boundary shows coefficients range where payoffs render investment into presentation of information non-profitable) and boundary between truthful strategy and false strategy.

Note there exist ranges of coefficient where the stable strategy is to lie even for honest producers and similarly to tell truth by liars.
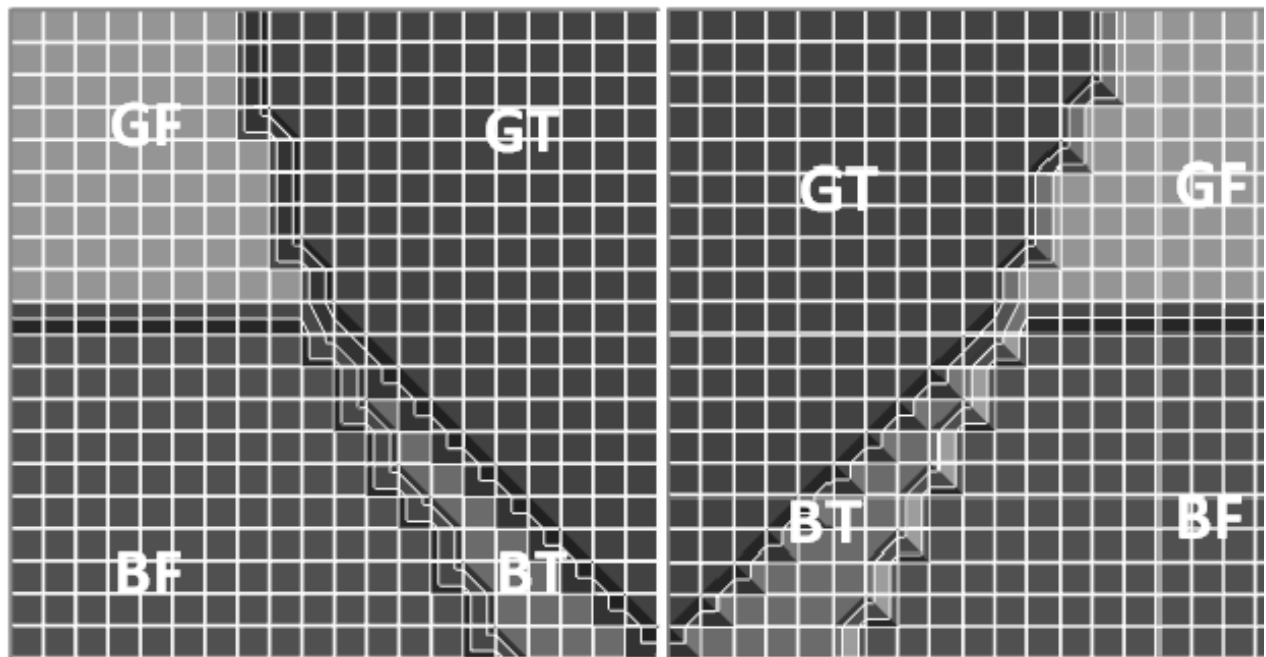

Figure 10. Stable strategy for CP-H (left) and CP-L (right) as parameter of α (x axis) and β (y axis) coefficients

9.15 The analysis shows that there exists a region (cone) of α and β where for both types of producers, the stable strategy is GT (this is the intersection of the two GT regions on the two sides of the figure), and another region, where the stable strategy is BF. In these two regions, studying phenomena such as the influence of expertise or reputation becomes impossible. The coefficient values chosen for the version of the Credibility Game studied in this paper do not lie in these regions and therefore, a mixture of strategies for the two types of CP can be expected to emerge from the simulations, depending on additional parameters.

Notes

[1]Research supported by the grant "Reconcile: Robust Online Credibility Evaluation of Web Content" from Switzerland through the Swiss Contribution to the enlarged European Union, http://reconcile.pjwstk.edu.pl

[2] By stability of a simulation scenario we mean that all of our simulation runs produced similar results for that scenario. We do not analyze stability in a game theory sense.

[3] http://digg.com

[4] http://reddit.com

[5] http://stackexchange.com

References

ABRAMCZUK, K., Adamska, P., Papaioannou, T., Wierzbicki, A., & Aberer, K. (2012). Game-theoretic Models of Web Credibility. *Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality '12).*

CHAMMAH, A., & Rapoport, A. M. (1965). *Prisoner's Dilemma*. U. of Michigan Press.

CHEVALIER, J., & Mayzlin, D. (2006, August). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354.

DELLAROCAS, C. (2005). Reputation mechanism design in online trading environments with pure moral hazard. *Info. Sys. Research*(16), 209–230.

DELLAROCAS, C. (2006, October). Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms. *Management Science, 52*(10), 1577–1593.

FOGG, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the

credibility of web sites?: a study with over 2,500 participants. *2003 conference on Designing for user experiences, DUX '03.* New York: ACM.

FRIEDMAN, E., Resnick, P., & Sami, R. (2007). Manipulation-resistant reputation systems. *Time*, strony 4–29.

GLAZER, A., & Rubinstein, J. A. (2006). A study in the pragmatics of persuasion: a game theoretical approach. *Theoretical Economics, 1*(4), 395–410.

GUTTMAN, J. M. (1992). The credibility game: Reputation and rational cooperation in a changing population. *Journal of Comparative Economics, 16*(4), 619–632.

HARGITTAI, E. (2002). Second-Level Digital Divide: Differences in People's Online Skills. *First Monday, 7*(4).

JAIN, S., Chen, Y., & Parkes, D. C. (2009). Designing incentives for online question and answer forums. *10th ACM conference on Electronic commerce.* New York: ACM.

JOHNSTON, R., V.A. Crooks, and J. Snyder. 2012. "I Didn't Even Know What I Was Looking For": Examining the Decision-Making Processes of Canadian Medical Tourists. *Globalization and Health* 8:23.

MCAFEE, A., & Ghosh, P. (2011). Incentivizing high-quality user-generated content. *20th international conference on World wide web, WWW'2011.* New York: ACM.

MELNIK, M. I., & Alm, J. (2002, September). Does a seller's ecommerce reputation matter? evidence from ebay auctions. *Journal of Industrial Economics, 50*(3), 337–49.

METZGER, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *J. Am. Soc. Inf. Sci. Technol.* (58), 2078–2091.

ORTEGA, F., Gonzalez-Barahona, J. M. & Robles, G. (2008). On the Inequality of Contributions to Wikipedia. *HICSS '08: Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (p./pp. 304), Washington, DC, USA: IEEE Computer Society.

PAPAIOANNOU, T., & Stamoulis, G. (2008). Achieving honest ratings with reputation-based fines in electronic markets. *INFOCOM'2008.*

RAINIE, L., & Kommers, N. (2002). *Use of the Internet at Major Life Moments.*

RESNICK, P., Zeckhauser, R. J., Swanson, J., & Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics, 9*(2), 79–101.

SAMUELSON, L. (1997). Evolutionary games and equilibrium selection. *MIT Press.*

SENECAL, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing, 80*(2), 159–169.

SHER, I. (2011). Credibility and determinism in a game of persuasion. *Games and Economic Behavior, 2*(71), 409–419.

TANAKA, Y., & Yamamoto, K. (2011). Enhancing credibility judgment of web search results. *CHI'11.* New York: ACM.

THOMPSON, N. (2003, June 23). More companies pay heed to their "word of mouse" reputation. *New York Times.*

VOSS, J. (2005). Measuring Wikipedia. *Proceedings International Conference of the International Society for Scientometrics and Informetrics: 10th.*

WENG, Y.-H., Kuo, K. N., Yang, C.-Y., Lo, H.-L., Shih, Y.-H., Chen, C., & Chiu, Y.-W. (2013, January). Increasing utilization of Internet-based resources following efforts to promote evidence-based medicine: a national study in Taiwan. *BMC Medical Informatics and Decision Making.*

WILSON, D., & Kreps, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*(27), 253–279.