

# Game-theoretic Models of Web Credibility

Thanasis G. Papaioannou, Karl Aberer  
Ecole Polytechnique Fédérale de Lausanne (EPFL)  
1015 Lausanne, Switzerland  
firstname.lastname@epfl.ch

Katarzyna Abramczuk  
Institute of Political Studies, Polish Academy of  
Sciences, Warsaw, Poland  
katarzyna.abramczuk@isppan.waw.pl

Paulina Adamska, Adam Wierzbicki  
Polish-Japanese Institute of Information  
Technology (PJIT), Warsaw, Poland  
{tiia,adamw}@pjwstk.edu.pl

## ABSTRACT

Research on Web credibility assessment can significantly benefit from new models that are better suited for evaluation and study of adversary strategies. Currently employed models lack several important aspects, such as the explicit modeling of Web content properties (e.g. presentation quality), the user economic incentives and assessment capabilities. In this paper, we introduce a new, game-theoretic model of credibility, referred to as the Credibility Game. We perform equilibrium and stability analysis of a simple variant of the game and then study it as a signaling game against naïve and expert information consumers. By a generic economic model of the player payoffs, we study, via simulation experiments, more complex variants of the Credibility Game and demonstrate the effect of consumer expertise and of the signal for credibility evaluation on the evolutionary stable strategies of the information producers and consumers.

## Categories and Subject Descriptors

H.3.5 [Online Information Services]: Commercial services, Web-based services; J.4 [Social and Behavioral Sciences]: Economics; I.2.6 [Artificial Intelligence]: Learning

## General Terms

Theory

## Keywords

incentives, signaling game, equilibrium, evolution

## 1. INTRODUCTION

Research on supporting credibility assessment of Web content requires the availability of benchmarks or evaluation methods. These evaluation methods can be of at least two

kinds: The first uses declarative or behavioral data acquired from real users. The second uses models of reality, either in a form that can be treated analytically or in simulations. Both methods have different advantages. While using real data or experiments on real users is the best empirical method of evaluation, this method has several limitations, such as scale and exploratory power. An important advantage of using models is the ability of testing diverse what-if scenarios, such as various adversary strategies against credibility evaluation methods. For example, imagine that a new feature of textual Web content is empirically found to improve the correctness of credibility evaluations by 30% when employed. Researchers are now faced with the following questions: what would be the impact of the improved credibility evaluation on the strategies of content producers, e.g. improve content honesty or presentation quality? Would these strategies of content producers be economically sustainable?

In this paper<sup>1</sup>, we propose a new model of information credibility based on a game-theoretic approach. The proposed model is more specialized, as compared to previous work [10, 12], and in our view better-suited for studying the strategies of rational stakeholders involved in the production and the credibility assessment of Web content. Our game-theoretic model fulfills the following requirements:

1. Use asymmetric roles for users who can produce or consume information.
2. Explicitly model the quality of information produced and consumed by users.
3. Model the preferences of information producers and consumers.
4. Allow to take into account the use of credibility evaluation methods and to study their impact on the model.
5. Model economic incentives of information producers and consumers.
6. Model diverse strategies of information producers and consumers.

In our simple, yet expressive, game, we model the encounter of two players in the Web: the *content producer* (CP) and *content consumer* (CC). (Of course, in a Web2.0 setting, a Web user can have both roles.) Diverse strategies may be specified for the CC for evaluating information

<sup>1</sup>This work was partially funded by the Swiss grant "Reconcile: Robust Online Credibility Evaluation of Web Content" through the Swiss contribution to the enlarged EU.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WebQuality '12, April 16, 2012, Lyon, France

Copyright 2012 ACM 978-1-4503-1237-0 ...\$10.00.

credibility, but the quality of information itself is also explicitly modeled. Web content credibility evaluation may depend on the content preferences of the CC and her ability to compare different content versions. For example, consumers with different subjectivity (or expertise) with respect to content credibility evaluation may be considered. The producer (CP) may also have preferences concerning the version of information that he wishes the CC to accept, e.g. a strategic (dishonest) producer may wish to trick the consumer, while an altruistic (honest) producer may wish to provide the most credible information. After all, in the real world, CP choices of content are influenced by monetary costs for content production and by monetary incentives (i.e. possible income) in case that a consumer accepts the content.

Content providers have different strategies regarding the truthfulness and the presentation quality of their Web content, depending on their honesty, while content consumers may accept or reject it, depending on their expertise. Initially, in the basic game form, we assume two CP types (of “high” and “low” credibility) and naïve CCs having a *finite* set of CP and CC strategies respectively, while players know the types of each other. We perform *equilibrium analysis* and study the *evolutionary dynamics* of this game. Then, we relax the assumption that players know the types of each other, study the encounter as a *signaling game* and find the perfect Bayesian equilibria against naïve and expert consumers. Subsequently, we explicitly model the economic incentives of content producers and consumers of different types for a *continuous* set of CP and CC strategies. This economic modeling is employed in our simulations to study the evolutionary dynamics by means of “learning” (e.g. aspiration and imitation) of different populations mixes of CP and CC types. Moreover, our simulation results verify our game-theoretic analysis.

Our model could incorporate diverse strategies of content producers and consumers, and different preferences or incentives. Presentation quality is explicitly present in the model, but, at the same time, it is expressively powerful to encapsulate a number of different Web content features that may affect the credibility evaluation. Also, our model provides the ability to study the impact of the various aspects, e.g. a 30% improvement in credibility assessment, on the strategies of content producers and consumers at equilibrium. Our model is still simple enough to be analytically tractable, and ease to simulate.

The structure of this paper is as follows: in the next section, we describe related work on game modeling of credibility and on features for assessing Web credibility. In Section 3, we introduce our basic version of the *Credibility Game*, explain how it achieves the requirements described above and perform equilibrium and stability analysis. In Section 4, we revisit the Credibility Game as a signaling one and find its assessment equilibrium; a signal for the consumers regarding the content quality is communicated by producers. Section 5 provides a generic economic modeling of player payoffs in credibility games. In Section 6, we study the evolution of the game under different the properties of the signal, different producer honesty and different consumer expertise. Finally, in Section 7, we conclude our work.

## 2. RELATED WORK

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 hidden quality [7], where adverse selection leads to the “market-of-lemons” effect, or to hidden type, where moral hazard [2] 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 [7, 2]. Such an approach would necessitate the existence of *credible* ratings for the Web pages subject to assessment. Although initial attempts towards this goal have started (e.g. Google+, plus.google.com), it will be long before enough adequate user feedback is 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.

There is an extended literature on game models for online reputation systems [2, 9], for a survey, see [3]. These games study the incentives of transacted parties in an encounter towards honesty [2] and truthful feedback [9]. [5] studies voluntary cooperation in case of labor-managed firms as a Prisoner’s Dilemma, i.e. players may independent decide to cooperate or defect, while defect is a dominant strategy. 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. 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.

In [6, 4], the issue of incentivizing high quality of user generated content is addressed with a game-theoretic modeling. In a generic model, [4] studies 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 approach for eliminating low-quality content based on a single negative rating. However, such an approach is mostly relevant for blog comments and reviews and could not be employed for arbitrary Web sites; it could easily lead to defamation of honest Web sites by rational competitors.

A game of persuasion is studied in [10, 12]. 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 from which the speaker chooses. A persuasion rule specifies which statements the listener finds persuasive. In [10], persuasion rules are studied 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. In [12], it is shown that concavity is the critical assumption for both aforementioned results: The listener’s utility function is a concave transformation of the speaker’s utility function. The model in [12] is broadened beyond binary listener’ decisions given the concavity assumption. 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.

Another related category of games concern arms races, e.g. data miner against adversary feature manipulation [1], Web spam filters against sophisticated spammers, etc. In [1], classifier chooses to utilize certain features based on their cost for classifying data, while adversary exerts different costly effort for manipulating different features in order for classifier to generate false negatives. Classifier’s best response strategy is derived given rational adversaries. However, the evolutionary dynamics of the game in [1] have not been studied, while different content producer and consumer strategies cannot be easily modeled, as opposed to our work.

Regarding Web credibility assessment, in [13], six measurable factors are identified, which are 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. Other methods that automatically assess credibility aggregate different features values such as [8]: 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.

### 3. BASIC MODEL OF CREDIBILITY

In this section, we introduce the basic form of our credibility model and then describe the basic version of the Credibility game. In order to realize Requirement 1, our model uses the concept of encounters between two different kinds of users (players): the Content Provider (CP) and Content Consumer (CC). We assume here that in an encounter, CP produces and CC consumes information. The context of the encounter is fixed, and part of it is the subject of the information that CP produces and CC consumes.

CP’s actions are a choice of variants  $\{I_i\}$  of the information  $I$  that CP can provide to CC. In order to fulfill Requirement 2, the model should incorporate some properties of  $I_i$ . There may be various ways of doing this, but in this paper, we shall focus on two information properties: truthfulness (TF) and looks (L) with  $TF \in [0, 1]$  and  $L \in [0, 1]$ . Information truthfulness depends on the honesty of its CP, while by information looks, we refer to the presentation quality of

the content. Following the work on persuasion games, we can model CC’s actions in the encounter as binary (CC can accept or reject the version of information produced by CP).

Requirement 3 will be met by our model if we are able to express the preferences of CC and CP regarding the version of information that CC accepts. The truthfulness (TF) property of the versions of information creates an objective correctness ordering of the versions of  $I$ :  $I_i \ll I_j$ . The preferences of CP can be modeled as follows. The looks property (L) of the versions of information models the investment that CP can make in the produced information. Increasing  $L$  can increase the likelihood that CC will accept the produced version of  $I$ . Yet, this investment is costly (increasing  $L$  increases the 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 depend on the TF property. We can have different types of producers with different preferences. An honest CP will prefer that CC would accept a truthful version of  $I$ : CP prefers that CC accepts  $I_i$  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.

CC’s preferences depend solely on TF. Note that CC does not need to know in advance what the TF of the version of information produced by CP will be. After CC accepts the content, she will “find out” (through experience) how high it ranks in the correctness ordering. CC may also use strategies that rely on her estimation about the produced version of  $I$ . This estimation may depend on both  $TF$  and  $L$  and may be imprecise, probabilistic, or subject to systematic error (for example, biased for higher versions of  $L$ ).

CC may also use CP’s reputation in her actions based on recommendations by previously encountered CCs. Moreover, if there are multiple CP players, then CC’s action may be a choice of the variants of  $I$  provided by the various CP players, or a rejection of all offered variants of  $I$ . Notice that if CC can provide recommendations, then we are changing the encounter, and CC becomes CP. The reputation of the recommendation producer will be a function of its strategy that is understood as both a strategy of choosing information to consume (when it was CC) and of providing information (the recommendation). Since reputation-based approaches have been studied in previous work, we will not focus on such an approach in this paper. Basing CC’s actions onto a multi-parametric estimation of the information truthfulness (having CP’s reputation as one of the parameter) is left for future work.

#### 3.1 The Basic Credibility Game

The basic Credibility game uses four versions of  $I$ :  $F$  (bad looking false),  $GF$  (good looking false),  $BT$  (bad looking true), and  $T$  (good looking true). The correctness ordering is:  $F = GF \ll T = BT$ . These versions are described by the following parameters  $TF$  and  $L$ :  $F:(TF = 0, L = 0)$ ;  $GF:(TF = 0, L = 1)$ ;  $BT:(TF = 1, L = 0)$ ;  $T:(TF = 1, L = 1)$ .

For simplicity, we consider two types of CP: the honest

CP (CP-H) and the dishonest CP (CP-L). The payoffs of CC, CP-L and CP-H model their preferences as described above. We denote as  $CP_{X,Y}$ ,  $CC_{X,Y}$  the expected payoffs of CP and CC respectively, when CP plays strategy  $X$  and CC plays strategy  $Y$ .

**Payoffs of CC.** CC is indifferent in case of rejection ( $R$ ):

$$CC_{F,R} = CC_{GF,R} = CC_{BT,R} = CC_{T,R} = 0$$

CC's payoffs in case of acceptance ( $A$ ) depend only on  $TF$ .

$$CC_{F,A} = CC_{GF,A} = -2 < CC_{BT,A} = CC_{T,A} = 2$$

**Payoffs of CP-L.** In case of rejection ( $R$ ), CP's payoffs depend inversely on  $L$  and increase with  $TF$ , i.e.

$$CP_{GF,R}^L = -3 < CP_{FR}^L = -2 < CP_{TR}^L = -1 < CP_{BTR}^L = 0.$$

In case of acceptance ( $A$ ),  $CP^L$ 's payoffs depend inversely on  $TF$  and on  $L$ , i.e.

$$CP_{TA}^L = 1 < CP_{BTA}^L = 2 < CP_{GFA}^L = 4 < CP_{FA}^L = 5.$$

**Payoffs of CP-H.** In case of rejection, CP-H has the same payoffs as CP-L. In case of acceptance, CP-L's payoffs depend inversely on  $L$  and increase with  $TF$ :

$$CP_{GFA}^H = 1 < CP_{FA}^H = 2 < CP_{TA}^H = 4 < CP_{BTA}^H = 5$$

		CC	
		A	R
CP-L	F	5	-2
	GF	4	-3
	BT	2	0
	T	1	-1

**Table 1: Basic game of CP-L vs. CC.**

		CC	
		A	R
CP-H	F	2	-2
	GF	1	-3
	BT	5	0
	T	4	-1

**Table 2: Basic game of CP-H vs. CC.**

### 3.2 Analysis of Basic Credibility Game

**For CP-L.** The game has zero equilibria in pure strategies and one Nash equilibrium in mixed strategies. First, observe that for CP-L the strategy  $GF$  is dominated by  $F$  and strategy  $T$  is dominated by  $BT$ . Then, assume that CP-L plays the strategy  $F$  with probability  $x$  and the strategy  $BT$  with probability  $1 - x$ . Also, assume that CC plays the strategy  $A$  with probability  $y$  and the strategy  $R$  with probability  $1 - y$ . Then, the unique Nash equilibrium is given by  $x = 1/2$  and  $y = 2/5$ .

In order to study the evolutionary stability, we take the replicator dynamics equations [11] to be 0, i.e.

$$\begin{aligned} \frac{\dot{x}}{x} &= 5y - 2(1 - y) - [x(5y - 2(1 - y)) + (1 - x)2y] = 0 \\ \frac{\dot{y}}{y} &= -2x + 2(1 - x) - [y(-2x + 2(1 - x)) + (1 - y)0] = 0 \end{aligned} \quad (1)$$

After some algebra, we obtain two solutions,  $x = 1$ ,  $y = 1$  and  $x = 1/2$ ,  $y = 2/5$ . However, the solution  $x = 1$ ,  $y = 1$  is just a fixed point and not evolutionary stable strategy (ESS), since it is not a Nash equilibrium; an ESS must also be a Nash equilibrium. Therefore, the only ESS of the game is  $x = 1/2$ ,  $y = 2/5$ .

**For CP-H.** The game has 1 equilibria in pure strategies and 0 Nash equilibrium in mixed strategies. Observe that for CP-H the strategy  $BT$  is dominant. Anticipating this, CC's best response is the strategy  $A$ . Therefore,  $(BT, A)$  is the subgame-perfect equilibrium of the game. Since  $(BT, A)$  is subgame-perfect equilibrium, it is also ESS.

## 4. CREDIBILITY SIGNALING GAME

The basic game assumes that the CP-type is observable by CC, while CC cannot deduce any belief on CP's strategy ( $TF, L$ ). Moreover, in the basic game, CP can select the content truthfulness regardless of her type. Here, we assume that the CP-type determines the content truthfulness, i.e. CP-H (resp. CP-L) produces truthful (resp. false) content, but it is hidden from CC. CC always prefers true from false content and has to decide to Accept( $A$ ) or Reject( $R$ ) the Web page information based on the observation of the signal jointly composed of the truthfulness  $TF$  and the presentation quality  $L$  of the information. This signal has an important role in our model. It can be used to fulfill Requirement 4. Improving the credibility judgments of information consumers can be modeled by increasing the role of  $TF$  in the signal. A perfect credibility judgment would be achieved if the signal depended only on  $TF$ . To this end, at the one extreme, naïve consumers mostly take into account the presentation quality  $L$  of the Web page instead of its truthfulness  $TF$ , while expert consumers will value mostly its truthfulness. In between these two extremes, several versions of the signal can be created that depend on both  $TF$  and  $L$ . The simplest strategy of CC towards accepting or not the information of the Web site could use a threshold for the signal value: if the signal is above the threshold, CC accepts the produced version of  $I$ .

For simplifying the analysis, we assume two types of information producers: honest (CP-H) that always produce truthful information and dishonest (or liars, CP-L) that always produce false information. However, producers may exert or not a costly effort to improve the presentation quality of the information. To this end, CP-H can produce information versions  $T$ ,  $BT$  by deciding to invest in presentation quality or not. On the other hand, CP-L can produce the information versions  $GF$ ,  $F$  if he invests on presentation quality or not respectively. It will also be possible to observe the effect of a change in the signal on the equilibrium analysis of the game and on the strategies of CC and CP.

CC is randomly matched with a CP of certain type. Given a population mix with a fraction  $r$  of honest providers, we assume that Nature selects with probability  $r$  that the CP has true information (i.e. type CP-H) and with probability  $1 - r$  that he has false information (i.e. type CP-L).

In this game, the payoffs of the players are those of the basic credibility game given in the Tables 1 or 2 for the cases that CC is matched with CP-L or CP-H respectively. Recall that the effort for high quality presentation subtracts 1 from the producer payoffs. The strategic form of this signaling game is depicted in Figure 1.

Assume that in the FALSE information set, CP plays Not Invest (i.e. bad looks,  $L = 0$ ) with probability  $x_1$ , while he invests (i.e. good looks,  $L = 1$ ) with probability  $1 - x_1$ . Also, in the TRUE information set, CP plays Not Invest ( $L = 0$ ) with probability  $x_2$ , while he invests ( $L = 1$ ) with probability  $1 - x_2$ . On the other hand, in the BAD LOOK information set, CC plays Accept with probability  $y_1$ , while,

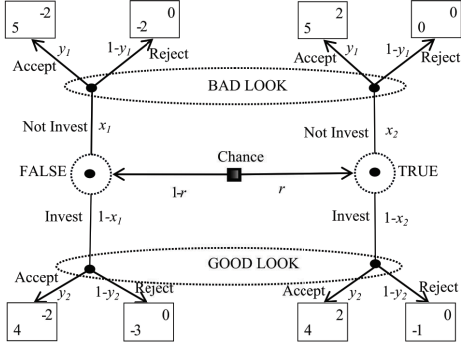


Figure 1: The signaling game of credibility.

in the GOOD LOOK information set, he plays Accept with probability  $y_2$ .

In the TRUE information set, CP-H should be indifferent in terms of expected payoffs between investing or not on high quality presentation, therefore, it should be for CP-H that

$$x_2(5y_1 + (1 - y_1)0) = (1 - x_2)(4y_2 - (1 - y_2)). \quad (2)$$

In the FALSE information set, CP-L should similarly be indifferent between investing or not on high quality presentation, and thus it should be for CP that

$$x_1(5y_1 - (1 - y_1)2) = (1 - x_1)(4y_2 - 3(1 - y_2)). \quad (3)$$

Regarding the CC strategies, we employ forward induction as follows: The unconditional probability that the game reaches the right node of the BAD LOOK information set is  $x_2r$ , while for the left node it is  $x_1(1 - r)$ . At the right node of BAD LOOK, CC would always play Accept, while at the left node of the same information set, he would always play Reject. Thus, after observing the BAD LOOK, the conditional probability that CC assigns of being to the right node should equal its probability  $y_1$  of playing Accept in this information set, i.e. it should be that

$$y_1 = \frac{x_2r}{x_2r + x_1(1 - r)}. \quad (4)$$

Reasoning similarly for the GOOD LOOK information set, the probability  $y_2$  of playing Accept is given by

$$y_2 = \frac{(1 - x_2)r}{(1 - x_2)r + (1 - x_1)r}. \quad (5)$$

The solution to the system of equations (2), (3), (4), (5) for  $r=1/2$  is given by:  $x_1=0.0283$ ,  $x_2=0.2176$ ,  $y_1=0.8846$ ,  $y_2=0.4460$ . This is an assessment (or perfect-Bayesian) equilibrium of this game.

We study the cases of completely naïve consumers and of those with some expertise separately below.

#### 4.1 Naïve Consumer

Here, we assume that CC is completely naïve and only considers whether the Web page is of high presentation quality or not for credibility evaluation. In other words, CC receives a signal that depends only on the  $L$  property of the versions of information  $I$ . Recall that the truthfulness of the information depends on whether the CC is matched with CP-H or CP-L. CC plays Accept with probability  $y_1 = 0$  in the BAD LOOK information set, while she plays Accept with probability  $y_2 = 1$  in the GOOD LOOK information

set. Anticipating this, it is dominant strategy for both CP-L and CP-H to invest in presentation quality against completely naïve consumers, i.e.  $x_1 = x_2 = 0$ .

#### 4.2 Consumer Expertise

In fact, the game model of Web credibility as a signaling game (see Figure 1) is very expressive and can be used to study the behavior of information producers and consumers for various mixes of CP-H and CP-L and for various expertise levels of the CC. If the CC is highly-skilled expert, he can correctly assess the credibility of the information and find the type of the CP regardless of the presentation quality of the information. In other words, whether in BAD LOOK or in GOOD LOOK, CC knows if he is at the right or the left node of the respective information set and plays accordingly. A highly-skilled expert CC would play Accept (A) when matched with CP-H and Reject (R) when matched with CP-L. Anticipating this, neither CP-H or CP-L has any incentive to invest in high presentation quality against an expert CC, i.e.  $x_1 = x_2 = 1$ .

On the other hand, if in the CC population consisted of a fraction  $f$  of highly-skilled experts, while the rest were totally naïve, CP would not know in advance the type of the CC that is matched with. Then, reasoning similarly to the equilibrium analysis above, CP-H would always play Invest (i.e.  $x_2 = 0$ ) if  $f < 0.8$ , otherwise Not Invest ( $x_2 = 1$ ); observe that, in the right node of the BAD LOOK information set, only expert consumers are expected to play Accept, while in the right node of the GOOD LOOK information set all consumers play Accept. Also, CP-L would always play Invest ( $x_1 = 0$ ) if  $f < 6/7$ , otherwise Not Invest ( $x_1 = 1$ ); here, observe that in the left node of the GOOD LOOK information set, only naïve consumers play Accept, while in the left node of the BAD LOOK information set, all consumers play Reject.

In general, there can be CC with various levels of expertise according to which CC assigns a conditional probability to encounter a CP-H or a CP-L given a certain presentation quality (i.e. GOOD LOOK or BAD LOOK information sets). These conditional probabilities determine  $y_1$ ,  $y_2$  with whom CC plays Accept in the BAD LOOK and the GOOD LOOK information sets respectively and the  $x_1$ ,  $x_2$  are then derived by equations (2) and (3).

#### 5. ECONOMIC MODEL OF CP PAYOFFS

In this section, we show how it is possible to realize Requirement 5. We do so by introducing a utility model for the content producer that depends on a continuous set of strategies regarding truthfulness  $TF$  and presentation quality  $L$  of the content, and it is consistent with the payoffs proposed in Section 3. The proposed utility model can be employed for economic modeling of information production in more generic cases. In this model, the gain of the CP depends only on  $TF$  (increasingly for CP-H and decreasingly for CP-L), and the cost of producing information depends increasingly on  $L$ , but decreasingly on  $TF$ . The latter property can be interpreted as it is more costly to manufacture a falsehood than to tell the truth.

The utility of a content producer is given by:

$$U(I) = kG(TF) - C(TF, L), \quad (6)$$

where  $k$  is the number of consumers that have accepted the CP's content,  $G(\cdot)$  is a function that describes the gain of

CP, and  $C(.,.)$  is a function that describes the cost of information production. This utility function models the fact that the marginal costs of producing information is 0: the CP only bears a fixed cost of production that does not depend on the number of consumers.

The gain of the content producer is given by:

$$G(TF) = \alpha TF + \beta, \quad (7)$$

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

The cost of the content producer is given by:

$$C(TF, L) = \gamma TF + \delta L + \epsilon, \quad (8)$$

where  $\gamma < 0$  models that increasing the truthfulness  $TF$  of the produced information reduces the cost of its production (the cost is determined by the cognitive effort for manufacturing a falsehood). Also,  $\delta > 0$  models that improving the looks increases the cost, while  $\epsilon > 0$  models a fixed cost component.

Consider the following exemplary parameters for CP-L:  $\alpha = -5, \beta = 7, \gamma = -2, \delta = 1, \epsilon = 2$ . The resulting utility function is given by:

$$U^L = \begin{cases} -3TF - L + 5, & k = 1 \\ 2TF - L - 1, & k = 0 \end{cases} \quad (9)$$

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

Also, consider the following exemplary parameters for CP-H:  $\alpha = 1, \beta = 4, \gamma = -2, \delta = 1, \epsilon = 2$ . The resulting utility function is given by:

$$U^H = \begin{cases} 3TF - L + 2, & k = 1 \\ 2TF - L - 1, & k = 0 \end{cases} \quad (10)$$

This utility function produces the payoffs for CP-H for the basic Credibility game. Notice that the cost functions for CP-H and CP-L do not differ.

## 6. EVALUATION

In this section, we experimentally analyze the evolution of the credibility signaling game under provider population mixes of varying honesty and under consumer population mixes of varying expertise. Moreover, we vary the transparency of the truthfulness signal to the consumers to assess cases where different presentation tricks can be employed by dishonest producers to mislead consumers towards information acceptance.

### 6.1 Simulation Model

Our simulation model builds upon and extends the model described in Section 4, while employing the generic economic payoff modeling of Section 5. CC's expertise is not binary and it is modeled as a standard deviation of signal that CC receives before making decision to Accept or Reject the information. It is assumed that the CP signal has a normal distribution. Its standard deviation for different consumers may equal 0.05, 0.083, 0.16, 0.33, 0.66 or 1 according to their expertise; of course, the larger the standard deviation, the more naïve the consumer. The mean of the signal is a weighted average of information truthfulness  $TF$  and presentation quality  $L$ . The weight of  $TF$  varies for different simulation experiments and it can be 0.1, 0.33, 0.66 or 0.9.

Recall that, the greater the  $TF$  weight, the "better" the signal reflects the information truthfulness. Since the signal that CC receives is not binary, there are more than two information sets in which CC has to make a decision, as compared to Section 4. CC's strategy is a certain threshold  $t \in [0, 1]$ , above which CC accepts the information and below which he rejects it.

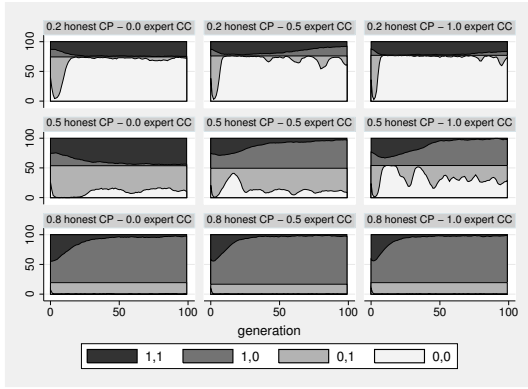
We assume a population of 100 producers and 1000 consumers. In the beginning of each simulation experiment, the types of producers (liar or honest) and consumers (expertise level) are determined based on the specific simulation parameters. The player types do not change throughout the game. A single game round starts with producers choosing whether to produce information that is good- or bad-looking according to their strategies, while truthfulness in all but the last experiment depends on the producer type. Then, each consumer randomly chooses one producer from the population and receives the signal according to the information properties ( $TF$  and  $L$ ) and the expertise of the consumer. Based on this signal and the acceptance threshold of the consumer, the consumer decides to accept or not the information. Every 100 rounds (i.e. learning period, referred to as *generation*) evolution takes place. It is applied separately to each type of CC and CP. Entities produce offsprings proportional to their payoff<sup>2</sup> and "die". The offspring inherit parental strategy. We also include mutation by assuming that, with 1% probability, a producer does not reproduce accurately; instead, its offspring chooses a random strategy. For consumers, we also assume that inheritance is not perfect upon mutation and the offspring chooses a threshold that equals parental threshold  $\pm 0.05$ . A single simulation run lasts for 100 generations. In the next subsection, we explore how the simulation results vary with three factors: share of liars in producers population, fraction of expertise in consumer population and weight of  $TF$  in the information signal.

## 6.2 Results

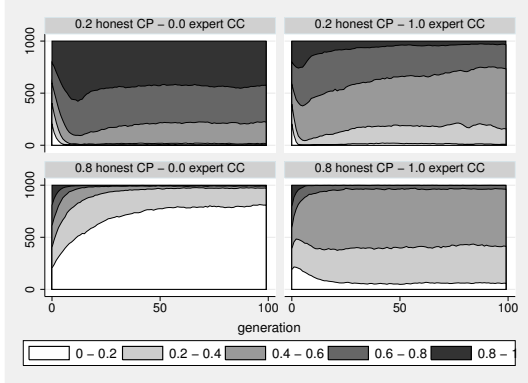
We first consider the case where the truthfulness of the Web page is a more powerful signal to the consumer than its presentation quality, i.e. signal truthfulness weight: 0.66, signal looks weight: 0.33. When honest content producers constitute the 80% of the producer population (top graphs of Figure 2(a)), honest CPs do not invest in high presentation quality, as opposed to dishonest ones, when facing either experts or naïve CCs. When the population of producers contains roughly 50% of honest ones (second line of graphs in Figure 2(a)), then producers exploit different behaviors when facing consumers of different expertise; honest and liar producers tend to invest in high presentation quality (i.e.  $L = 1$ ) when facing mostly naïve consumers, while they do not invest in presentation quality (i.e.  $L = 0$ ) when facing mostly experts, as predicted by the analysis of Section 4.

The evolution of the consumer acceptance thresholds for different producer populations is depicted in Figure 2(b). Naïve consumers employ extremely high (resp. low) threshold when facing mostly liar (resp. honest) producers. Expert consumers always employ moderately high acceptance thresholds that tend to be higher when the producer population contains more liars.

<sup>2</sup>A constant is added to payoffs to assure that they are not negative. Stochastic universal sampling is used to minimize genetic drift.



(a)

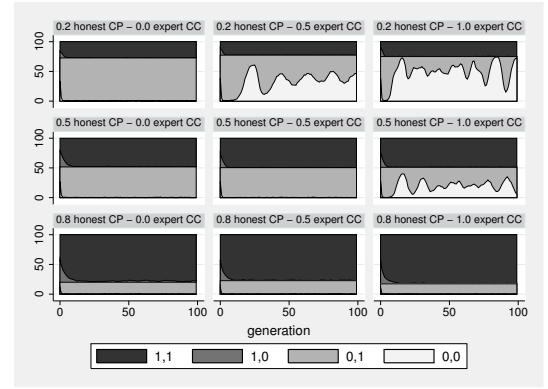


(b)

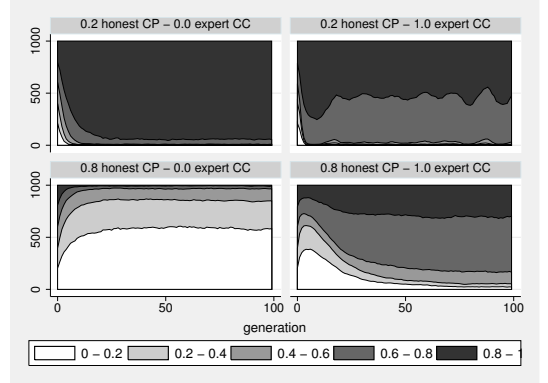
**Figure 2: Truthfulness is a more powerful signal than looks.** (a) Producers strategy  $L$  evolution regarding presentation quality when facing consumers with different expertise. (b) Evolution of consumer acceptance thresholds when facing different CP mixes.

If the consumers give more significance to the presentation of the page (i.e. signal truthfulness weight: 0.33, signal looks weight: 0.66), then the evolution of producers regarding their investment strategy  $L$  in high quality presentation is depicted in Figure 3(a). As can be observed therein, honest producers always invest in high presentation quality, in order to be differentiable from dishonest ones. The dishonest producers invest less in presentation quality as consumers become more expert and as the producer population contains more liars, since i) expert CCs cannot be easily fooled by presentation quality to assess false content as credible and ii) when the fraction of honest producers in the CP population is low, then the acceptance thresholds employed by the CCs tend to be higher (see Figure 3(b)). As depicted in Figure 3(b), when presentation signal is stronger than truthfulness, acceptance rates of consumers are higher than those when truthfulness provides a stronger credibility signal than looks (see Figure 2(b)).

The impact of the strength of the truthfulness signal for credibility assessment on the  $L$  strategy of producers is depicted in Figure 4. Here, the producer population contains 50% honest ones, while the consumer population contains 50% experts. However, as expected, if the presentation quality signal is much appreciated by the consumers (top left graph in Figure 4), then all consumers behave as naïve ones, and thus all producers invest in presentation quality. On the



(a)

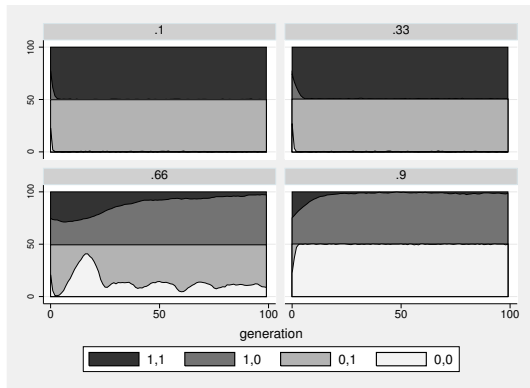


(b)

**Figure 3: Looks signal is a more powerful than truthfulness.** (a) Producers strategy evolution regarding presentation quality when facing consumers with different expertise. (b) Evolution of consumer acceptance thresholds when facing different CP mixes.

other hand, if the information truthfulness signal is strong for the consumers, then all consumers behave as experts, and thus all producers do not invest in presentation quality. These results totally agree with the analysis of the signaling game in Section 4. Specifically, it was expected that for highly-skilled expert consumers it would be at equilibrium that  $x_1 = x_2 = 0$ , while for totally naïve ones it would be  $x_1 = x_2 = 1$  at equilibrium. Our experimental results for the other weights of the information truthfulness in the signal can still be explained based on the analysis of Section 4.

Finally, we allow producers of both types (i.e. initially 50% honest and 50% liars) to modify their truthfulness strategy  $TF$  along with their strategy  $L$  for presentation quality according to their economic fitness given their respective payoffs in Tables 1, 2. Figure 5 depicts producer strategy evolution under different truthfulness signal strengths and different consumer expertise. As expected, when presentation quality signal is strong (i.e.  $TF$  weight: 0.33), producers of both types invest in high presentation quality ( $L = 1$ ), while keeping their respective truthfulness strategies unaltered ( $TF = 1$  for honest and  $TF = 0$  for dishonest) when facing naïve consumers. As consumer expertise rises or truthfulness signal strength increases, all producers invest less in presentation quality, while dishonest producers become truthful!



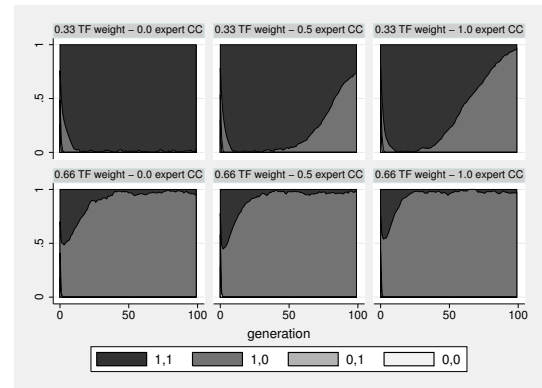
**Figure 4: Producers strategy evolution regarding presentation quality when truthfulness signal has different weights (0.1, 0.33, 0.66, 0.9) for consumers.**

## 7. CONCLUSION

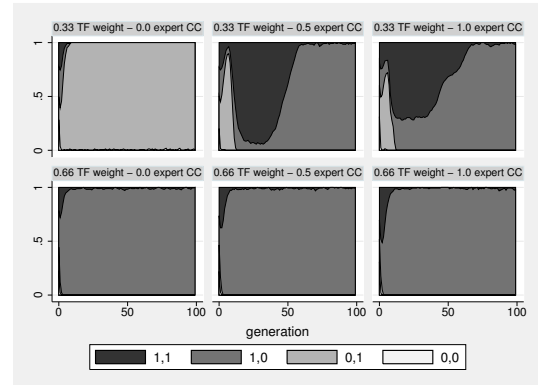
In this paper, we have proposed a new game model of Web content credibility. The model incorporates incentives and strategies of both producers and consumers of Web content. Using the proposed Web credibility game, it is possible to study the strategy evolution of populations of Web content producers and consumers with different honesty and expertise respectively. Therefore, it can be an important tool for modeling the strategy dynamics regarding the credibility of Web content. Also, presentation quality has been considered in our model as part of a signal to the content consumer. However, our game model is generic enough to consider arbitrary signaling means from the content producer to consumer. In the analysis and the simulation of the Credibility Game, we have considered simple strategies for both producers and consumers. As a future work, research on credibility assessment support should consider more sophisticated strategies. Moreover, incorporating user feedback on Web credibility is a natural extension to the model, which would add a certain consumer belief on the Web content producer type.

## 8. REFERENCES

- [1] N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma. Adversarial classification. In *Proc. of KDD*, 2004.
- [2] C. Dellarocas. Reputation mechanism design in online trading environments with pure moral hazard. *Info. Sys. Research*, 16:209–230, June 2005.
- [3] E. Friedman, P. Resnick, and R. Sami. Manipulation-resistant reputation systems. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, pages 4–29. Cambridge University Press, 2007.
- [4] A. Ghosh and P. McAfee. Incentivizing high-quality user-generated content. In *Proc. of WWW '11*, Hyderabad, India, 2011.
- [5] J. M. Guttman. The credibility game: Reputation and rational cooperation in a changing population. *Journal of Comparative Economics*, 16(4):619 – 632, 1992.
- [6] S. Jain, Y. Chen, and D. C. Parkes. Designing incentives for online question and answer forums. In *Proc. of EC '09*, Stanford, California, USA, 2009.
- [7] D. Kreps and R. Wilson. Reputation and imperfect



(a)



(b)

**Figure 5: Evolution of strategies (TF, L) of (a) honest and (b) dishonest producers under different strengths of truthfulness signal and different consumer expertise.**

information. *Journal of Economic Theory*, 27:253–279, 1982.

- [8] M. J. Metzger. 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, November 2007.
- [9] T. G. Papaioannou and G. D. Stamoulis. Achieving honest ratings with reputation-based fines in electronic markets. In *Proc. of INFOCOM '08*, Phoenix, AZ, USA, 2008.
- [10] A. Rubinstein and J. Glazer. A study in the pragmatics of persuasion: a game theoretical approach. *Theoretical Economics*, 1(4):395–410, December 2006.
- [11] L. Samuelson. *Evolutionary games and equilibrium selection*. MIT Press, Cambridge, Mass., USA, 1997.
- [12] I. Sher. Credibility and determinism in a game of persuasion. *Games and Economic Behavior*, 71(2):409–419, March 2011.
- [13] Y. Yamamoto and K. Tanaka. Enhancing credibility judgment of web search results. In *Proc. of CHI'11*, New York, NY, USA, 2011.