FLORE
Repository istituzionale dell'Università degli Studi di Firenze

Reputation matters the most: The reputation inertia effect

(Article begins on next page)

24 June 2024

WILEY

Reputation matters the most: The reputation inertia effect

Mirko Duradoni[1] ⬤ | Giorgio Gronchi[2] | Leonardo Bocchi[1] | Andrea Guazzini[3]

[1]Department of Information Engineering, University of Florence, Florence, Italy

[2]Department of Neuroscience, Psychology, Drug Research and Child's Health, University of Florence, Florence, Italy

[3]Department of Education, Languages, Intercultures, Literatures and Psychology (Psychology Section), University of Florence, Florence, Italy

Correspondence
Mirko Duradoni, Department of Information Engineering, University of Florence, via S. Marta 3, 50139, Florence, Italy.
Email: mirko.duradoni@unifi.it

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1002/hbe2.170.

Abstract

"Reputation systems" are widely used in a high number of web-based services to enhance cooperation among users, as well as to ensure they function well. However, the acquired reputation within such systems does not always reflect people's actual behavior. Consequently, this bias can reduce the effectiveness and robustness of a web-based system. The present study investigates the mechanisms with which reputation is built in an online multiplayer game. The reputation, once acquired, seems to be maintained over time (i.e., reputation inertia effect) despite the actual behavior of its owner. Moreover, if the players are asked to pay to suggest to the other players, the reputation inertia effect decreases. Nevertheless, even if reduced in frequency, "reputation inertia" persists under this condition.

KEYWORDS

on-line social influence, reputation dynamics, web psychology

1 | INTRODUCTION

In today's world, having a good reputation confers undoubtedly some advantages. Companies, firms, and freelancers know it well and commit energy and resources to reputation management practices (Dolle, 2014; Hutton, Goodman, Alexander, & Genest, 2001; Wæraas & Byrkjeflot, 2012). Such costs would not be expended without the belief that having a good reputation entails considerable advantages. Besides, the existence of such practices would not make sense if the reputation were not self-preserving to some extent over time (Van Der Heide, Johnson, & Vang, 2013).

The expansion of communication possibilities introduced by information and communication technologies (ICTs) has facilitated the development and the proliferation of systems based on online feedbacks (Dellarocas, 2003). The literature already explored trust and reputation construct, and refined method of assessments designed and validated (Bidgoly & Ladani, 2016; Chiregi & Navimipour, 2016).

However, despite many online services relying on reputation systems for their functioning, our knowledge is still limited about how reputation is attributed and its time course inside the cyber world, especially during the first stages of interaction, and with partial or incomplete information.

Scientific studies confirm the benefits of having a good reputation (Nowak & Sigmund, 1998; Ohtsuki & Iwasa, 2004; Sommerfeld, Krambeck, Semmann, & Milinski, 2007). In fact, by using the indirect reciprocity mechanisms offered by reputation, an individual can minimize the risks of being cheated. However, the indirect reciprocity mechanisms offered by reputation generally assume subjects' rationality as an axiom (i.e., giving positive feedback for those who have helped me and negative feedback for those who have harmed me). At the contrary, experimental evidence shows how, through their behavior, humans often violate the principle of rationality by using different behavioral and decision-making rules (e.g., social norms, heuristics; Biel & Thøgersen, 2007; Rand et al., 2014).

Therefore, it is of fundamental interest to understand whether humans adhere to the principle of rationality in attributing a reputation or whether they rely on different rules and norms. Interestingly, Semmann, Krambeck, and Milinski (2005) study has suggested that humans consider the past behavior of others (i.e., their reputation) more than their direct interaction with these partners. Reputation was able to exert a social influence in directing rewards and to overcome individuals' personal experience.

Despite the undoubted interest of this study, we do not know if the social influence of reputation (i.e., rewarding those who have good reputation regardless of their own experience) may also occur in

relation to informative behavior, and within environments in which individuals present conflicting interests, and have no personal incentive to provide feedback or evaluation (e.g., e-commerce sites). Moreover, in Semmann et al. (2005) public good game, the reputation was strictly bounded to the behavior (historical log of the decision to cooperate for the public good) and was not possible to let the reputation evolve on its own (e.g., build it based on individuals' feedbacks).

At this point, it becomes crucial to understand whether this persistence of the reputation remains when passing through a conceptualization of the reputation as the strict transposition of one's partner behavior, to an assignment conferred by others which may be more or less tied to a specific type of action (e.g., cooperation). Besides, social psychology studies have denoted that groups do not always allocate status fairly (Schneider & Cook, 1995), and this could lead to some "irrational" behavior, both in reputation building and in reputation maintenance processes. For instance, give positive feedback to a partner that did not cooperate but have a high reputation status. For clarity, from now on we will refer to a "rational agent" as that agent that rewards (positive feedback) his partner when he receives an advantage and punishes (negative feedback) him when he gets damage from him, disregarding his reputation.

For this reason, particular attention should be paid to those factors capable of bringing "rationality" back to the reputation building process, in all those situations in which the reputation evolves in a way that is too disconnected from the behavior which it should serve.

In order to investigate the "ecological reputation dynamics" in virtual environments, unlike the setting proposed by Semmann et al. (2005) in which the reputation served as an honest indicator of past cooperative conducts, we adopted a "widespread reputation building system" (i.e., in which reputation is built from the feedback of other individuals and not on the basis of the actual behavior). In this way, we determined in our system a higher degree of uncertainty and a more ecological measure of the subjects' reputation dynamics.

To study a possible solution to the "irrationality bias" affecting the reputation dynamics is required to consider the Costly Signaling Theory used in economics, evolutionary biology, and evolutionary psychology, that presents other possible ways of building and managing reputation (Barclay, 2011). For instance, when individuals pay a cost to help, they receive some benefits, including a reputational gain. People who pay for a prosocial act are seen as more trustworthy (Albert, Güth, Kirchler, & Maciejovsky, 2007; Barclay, 2006), and the payment conveys an informational value about the giver. Therefore, excited by the payment of others, the cognitive process involved in the construction of another's reputation appears becoming more data-driven (i.e., more focus is paid on the social partner's current behavior), so promising to be effective in reducing possible schema-driven biases. As a consequence, we introduced in our study even the opportunity to pay to have the possibility to advise another player, a very disadvantaging condition within the game, to study the effect of such condition on the social partners' behavior, and on the reputation dynamics itself.

Given the tendency of humans to use the internalized rules of conduct in new contexts and issues that have a certain degree of similarity (Rand et al., 2014), the following are the main hypotheses of the present study about the reputation dynamics within virtual environments:

H1 *Changes in reputation are affected by the level of reputation already achieved.*

H2 *Good reputation subjects tend to attract other positive feedback, regardless of personal/actual "experience." Conversely, bad reputation subjects are more likely to attract negative evaluations.*

H3 *Once acquired, reputation tends to be maintained/enhanced over time (i.e., it demonstrates inertia) in a way disconnected, at least in part, from the actual dynamics of the interactions.*

H4 *Paying a personal cost to provide information to others reduces the ability of the reputation to maintain itself in case of unreliable behaviors.*

To verify our hypotheses, we developed a social dilemma game called the bargaining game. It involved a widespread feedback system among the players in a competitive scenario. In some circumstances (i.e., game sessions), to evaluate other players' behaviors, an individual had to pay a personal cost, while in others did not. For further details about the game, please refer to the game section.

2 | METHODS

2.1 | Participants

The research was conducted following the guidelines for the ethical treatment of human participants of the Italian Psychological Association (i.e., AIP). The participants were recruited through a completely voluntary census and had no monetary incentives to take part in the experiments. All participants (or their legal guardians) signed an informed consent form and could withdraw from the experimental session at any time.

Overall, 203 participants (121 females) took part in our experiments. The sample size for our work has been determined using the work of Semmann et al. (2005) as the reference point. A brief presentation of the various samples will be described here, while their game-related descriptive statistics are presented in the results section. Seventy-seven adolescent volunteers (36 females) with an average age of 16 (*SD* 1.28) were recruited and carried out the experiment entirely in the Payment Off condition. Also, 36 adult volunteers (19 females) with an average age of 21 (*SD* 1.88) completed the experiment in the Payment Off condition. Ninety adult volunteers (66 females) with an average age of 22 (*SD* 3.45) underwent our experiment in the Payment On condition.

2.2 | The bargaining game

The game consisted of 45 independent rounds in which a Donor interacted with a Receiver and an Observer. Participants were anonymous

and identified only through nicknames. Participants played in groups of six, and each participant played all the roles of the game 16 times in a sequence randomly determined by a computer program. This game-dynamic resulted in each participant interacting three times in each role with each other group member. At the beginning of the game, each player was endowed with three kinds of resources, labeled Gold, Power, and Happiness, which were functionally equivalent. Among these resources, one was set equal to 50 units, and the other two were set at a minimal level of 5 units each. The software randomly distributed resources at the beginning of the session, and the player with the highest amount of the minimum resource at the end of the game was the winner. For the sake of clarity, we specify that for each player the type of the minimum resource dynamically varied during the game session according to his actions. In other words, if a player started with 50 Gold, 5 Power, and 5 Happiness and ended up with 10 Gold, 15 Power, and 20 Happiness, his final score is defined by the Gold resource disregarding the fact that was the maximum resource at the beginning. The players could see both their score and those of their opponents for the whole duration of the game. The players in each role had different tasks and goals (see Table 1).

The Donor's task was to make an offer to and a request of the Receiver. The Donor offered a given amount of his greatest resource, among the three at her disposal, and in return asked for a certain amount of the Receiver's minimum resources. Actual quantities were adjusted using sliders. The Receiver could only see the amount and type of the resource offered by the Donor but was unaware of what and how much the Donor had asked in return. The Receiver could "accept" or "reject" the Donor's deal right away, or could request the Observer's suggestion (by clicking on the "ask suggestion" button).

The Observer had the opportunity to evaluate the Donor's offer and request, knowing both the amount and the type of resources involved in the deal. The Observer could provide a hint to the Receiver, clicking on the button "suggest to accept," "no hint" or "suggest refusing." In the Payment On condition, providing a suggestion (i.e., to accept or to refuse) determined a payment equal to 1 of the Observer's highest resource, while selecting "no hint" meant that the Observer did not undergo any cost. Instead, in the Payment Off condition, none of the Observer's available action was charged a fee. The Observer had 10 s to make her choice.

TABLE 1 Roles

Role recap	
Roles	**Actions**
Donor	Offers his/her maximum resource and asks the *Receiver* his/her minimum resource.
Observer	Makes suggestions to the *Receiver* about the *Donor*'s offer and can receive a like or a dislike from the *Receiver*.
Receiver	Accepts or declines the *Donor*'s deal with no additional information or asks for the *Observer*'s suggestion. Can feedback on the *Observer*'s suggestion with a like/dislike.

Note: Summary of the actions to fulfill within the game for each role.

To decide whether to ask for the Observer's suggestion, the Receiver also had access to the rating (i.e., the number of likes and dislikes accumulated) of the Observer. When the game started all the Observers had a neutral reputation score (i.e., 0), and at each time step the Observer's reputation is updated Equations (1) and (2), where L_O^{t+1} and D_O^{t+1} represent respectively the number of likes and dislikes (i.e., L_R and D_R) accumulated by the Observer from the Receivers, before the time $t + 1$.

Observer's like equation:

$$L_O^{t+1} = \sum_{t^*=1}^{t} L_R^{t^*} \tag{1}$$

Observer's dislike equation:

$$D_O^{t+1} = \sum_{t^*=1}^{t} D_R^{t^*} \tag{2}$$

We did that to better simulate and study the reputation evolution in the early stages of a virtual setting with no prior information about its users. Once the offer is accepted or refused, the Receiver becomes aware of the Donor's request, and the resources are transferred. At this point, if the Receiver had asked for a suggestion, he/she would have the opportunity to give a like or a dislike to the Observer. We specify that even if the Receiver asked for a suggestion and obtained a "no hint" from the Observer, the latter still resulted eligible by our system to be evaluated by Receivers. Observers were not aware of the single evaluations received, nor of their overall reputation, but were informed at the very beginning that the Receivers would judge them and that the Receiver's feedback actions would determinate their reputation within the game. The Receiver had 18 s to make his decisions.

For all the roles, if a decision was not made within the available time frame, default options were set by the computer. In general, for each role, none of the possible actions were "externally" incentivized. For instance, provide reliable and coherent feedback was neither reward nor punished by our system. Even winning the game did not involve monetary rewards or prizes. In the Payment On condition, the only action to provide a suggestion was charged by a fee (on game resources) and thus disincentivized.

The bargaining game was developed as a multiplayer virtual game implemented through Google Apps, using the Google Script programming language.

2.3 | Surveys

2.3.1 | Sociodemographic survey

Participants were profiled according to their gender and age.

2.3.2 | Five-factor Adjective Short Test

Developed by Giannini, Pannocchia, Grotto, and Gori (2012) the Five-Factor Adjective Short Test (5-FasT) investigates the five-factor model

of personality traits: neuroticism, surgency, agreeableness, closeness, and conscientiousness. Using a 5-point Likert scale (ranging from "*not at all*" to "*very much*"), participants had to indicate how much the 26 adjectives described their personality (example items: anxious, active, calm, closed, confused, brave, distant).

2.3.3 | Self-efficacy Scale

Developed by Jerusalem and Schwarzer (1992), this scale investigated the perception of self-efficacy of the participants through 10 items. For each assertion, participants had to indicate their level of agreement through a 4-point Likert scale (ranging from "*not true at all* "to "*totally true*"). Examples of items are: "I can always manage to solve difficult problems if I try hard enough," "I am confident that I could deal efficiently with unexpected events."

2.3.4 | Classroom Community Scale

Developed by Rovai, Wighting, and Lucking (2004), this scale examined the sense of community about the participants' reference network. The scale consisted of 20 items and two subscales (social community and community learning); 10 items measured each of the subscales. For each statement contained, participants had to indicate their degree of agreement using a 5-point Likert scale (ranging from "*strongly disagree*" to "*strongly agree*"). Examples of items are: "I feel that I can rely on others in this course" (social community), "I feel that this course does not promote a desire to learn" (learning community).

2.4 | Procedure

The experimental sessions that involved adolescents were conducted in the computer lab inside the high school. The experiments concerning the adults were carried out in the computer lab of the Faculty of Psychology. Upon their arrival, the experimenter seated the participants at their designated computers and gave them a brief speech about the fact that their anonymity was assured. Moreover, to preserve the player's anonymity, all participants were separated by partitions. After providing the necessary demographic information (age, gender, years of education) and completing the psychological survey, the participants received instructions about the game that were read aloud and shown on the participants' monitors.

2.5 | Data analysis

In the first step, we verified the preconditions necessary for the inferential analyses on the experiment's data. For the continuous observables that were under investigation, the normality of the distribution was assessed through the analysis of asymmetry and kurtosis values. Then, due to the repeated measures structure of the experimental data, the inferential analyses were conducted using a general linear mixed model (GLMM) approach (McCulloch & Neuhaus, 2001).

3 | RESULTS

3.1 | Descriptive statistics

Table 2a reports the descriptive statistics for the game-related variables, already divided according to the sample type and game condition played. The descriptive statistics for the psychological and psychosocial observables are presented in Table 2b.

3.2 | Evaluation dynamics: How reputation is "made"

To better understand how the reputation was built and handled within our game (and so test our first hypothesis), we focused our attention on the feedback actions (i.e., give a like or a dislike) of the Receivers. Indeed, it was through the feedback that the Observer's reputation was built and shown to other players as the difference between the number of likes received minus the number of dislikes got in the Observer role. First, we investigated age-related differences regarding our observables through generalized linear mixed models.

As we can gather from Table 3, adolescents provided less frequently feedback (i.e., both likes and dislikes) and achieved on average

TABLE 2a Game observables descriptive statistics

	Samples		
Variables	**Adolescents payment off Mean (*SD*)**	**Adults payment off Mean (*SD*)**	**Adults payment on Mean (*SD*)**
Like	2.11 (1.13)	4.00 (2.42)	3.82 (1.56)
Dislike	2.28 (1.38)	2.82 (1.85)	3.30 (1.27)
Reputation	−1.65 (1.18)	1.23 (2.74)	0.53 (1.80)
Goodness of suggestion[a]	65.5%	66.7%	70.1%

Note: Game observables descriptive statistics for each sample involved.
[a]Percentage of good suggestions.

TABLE 2b Psychological and psychosocial descriptive statistics

	Samples		
Variables	**Adolescents payment off Mean (*SD*)**	**Adults payment off Mean (*SD*)**	**Adults payment on Mean (*SD*)**
Neuroticism	7.18 (3.78)	8.06 (3.77)	8.33 (4.10)
Surgency	11.17 (3.68)	10.44 (3.51)	10.41 (2.90)
Agreeableness	10.55 (3.90)	12.36 (3.07)	11.96 (3.29)
Closeness	5.31 (3.76)	5.75 (4.53)	5.42 (3.64)
Conscientiousness	7.87 (3.98)	8.33 (4.50)	9.61 (4.05)
Self-efficacy	17.52 (4.23)	19.31 (4.46)	18.33 (4.03)
Sense of community	22.04 (5.40)	26.31 (4.02)	22.18 (4.67)

Note: Psychological and psychosocial descriptive statistics for each sample involved.

TABLE 3 Generalized linear mixed models

Game observables age differences			
Variables	**F**	**Coefficient (β)**	**Student *t***
Like	182.702***	−2.062	−13.517***
Dislike	15.389***	−0.557	−3.923***
Reputation	105.580***	−1.505	10.275***
Goodness of suggestion[a]	0.158 ns	0.050	0.397 ns

Note: Game observables differences between adolescents and adults.
Abbreviation: ns, not significant.
[a]Percentage of good suggestions.
***$p < .001$.

TABLE 4 Generalized linear mixed models (GLMM)

GLMM best model like			
	Model precision	**Akaike**	**F**	**Df-1(2)**
---	---	---	---	---
Best model	77.1%	64.246	50.27***	6 (752)
Fixed effects				
Factor		**F**		**Df-1(2)**
Reputation		159.74***		1 (752)
Goodness of suggestion		80.14***		2 (752)
Payment[b] goodness of suggestion		3.96***		3 (752)
Parameter		**Coefficient (β)**		**Student *t***
Reputation(−)		−2.117		−12.64***
Goodness of suggestion(−)		−1.878		−6.65***
Goodness of suggestion(0)[a]		−2.813		−10.38***
Payment(1)[b] goodness of suggestion(0)[a]		−1.064		−3.05***

Note: Factors that influence the feedback behavior of the Receivers.
[a]Suggestion not present.
[b]Underlines the interaction between the variables.
***$p < .001$.

a lower reputation despite showing a similar suggestion behavior compared to adults. Given this evidence, we performed further generalized linear mixed models that considered as parameters the reputation score of the Observer and the goodness of the suggestion given by the Observer (defined as it follows: The Observer provided a "good" suggestion when he suggested to accept a deal when the Donor's offer was greater than or equal to his request and to refuse if that difference was lower than 0). Otherwise, the Observer's suggestion was classified as a "bad" suggestion), the genders of the participants playing as Receivers as well as their ages (i.e., adolescents and young adults). Furthermore, we took into account for the young adults' sample (adolescents played only one game scenario) the two different game settings offered by the presence or the absence of a

TABLE 5 Generalized linear mixed models

GLMM best model "reputation inertia"			
	Model precision	**Akaike**	**F**	**Df-1(2)**
---	---	---	---	---
Best model	75.0%	49.38	5.44***	6 (565)
Fixed effects				
Factor		**F**		**Df-1(2)**
Reputation[a] goodness of suggestion		27.32***		1 (565)
Parameter		**Coefficient (β)**		**Student *t***
Payment(0)[a] reputation(−)		−1.157		−2.89***
Payment(1)[a] reputation(−)		−1.486		−2.66***
Payment(0)[a] goodness of suggestion(−)		−1.918		−3.87***
Payment(1)[a] goodness of suggestion(−)		−1.517		−3.23***
Goodness of suggestion(−)[a] reputation(−)		3.326		5.23***

Note: Factors that promote the use of the reputation criterion across both conditions. (0): Payment not present.
Abbreviation: GLMM, generalized linear mixed model.
[a]Underlines the interaction between the variables.
***$p < .001$.

costly transmission of information for the Observer. The final model is reported in Table 4.

Age and gender did not appear to affect the feedback behavior of the Receiver either directly or through interaction effects in any of the subsequent models. In particular, the fact that adolescents gave less feedback did not appear to affect the way they give them, which was similar to the adults' behavior. Interestingly, only two factors contributed to forming the reputation of each participant in the Observer role: the goodness of the suggestion and the level of the reputation achieved. In other words, good suggestions and positive reputations more frequently attracted positive feedback from the others. Moreover, as we could appreciate from the standardized β in the Receivers' decision making, the reputation level of the partner seemed to outweigh the goodness of the hint received, thus supporting even more the H1 claim. Besides, refraining from providing a suggestion usually led more frequently to negative feedback. Furthermore, the game setting (i.e., Payment On/Off) seemed to influence the construction of the Observer's reputation marginally. Indeed, in the Payment On condition, the Observers who did not provide a hint to the Receivers were evaluated even worse.

3.3 | Evaluation coherence: How reputation alters decision making

As we have seen, reputation seemed a crucial factor in building and determining itself. Thus, we investigated whether this tendency to feedback to our social partners by relying on their previously acquired
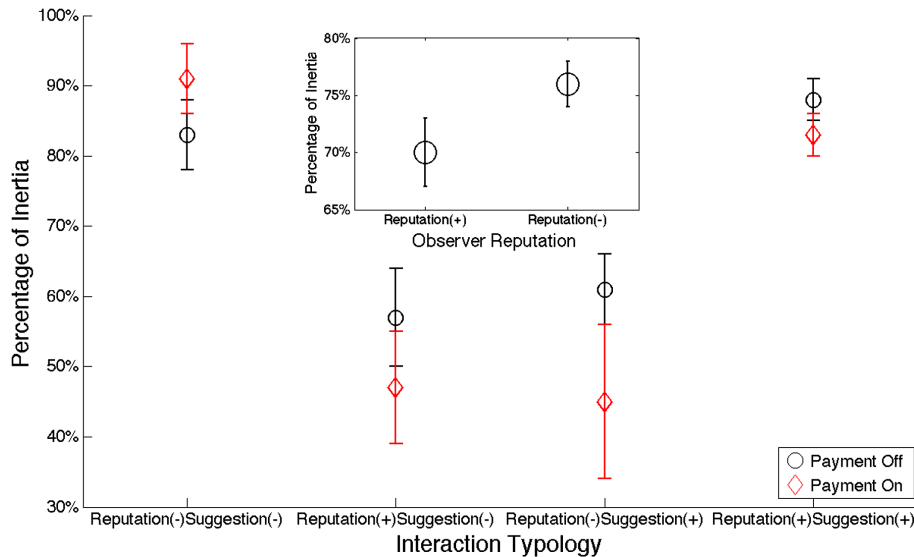
FIGURE 1 In the figure, the percentage of "inertia" is presented with respect of the experimental condition related to the payment (i.e., red and black bars), and the four possible "interaction typologies" (i.e., the combination between the Observer's reputation and the fairness of her suggestion. The error bars report the standard error of the mean. The insert figure highlights the relation between inertia and the Observers' reputation

reputation is influenced by sociodemographic, psychological, or game-related factors. The results of our GLMM between the two game settings (i.e., Payment On/Off) are reported in Table 5.

As we could appreciate from the model, neither the sociodemographic nor the psychological variables entertained a significant relationship with the tendency to use the reputational information to provide feedback. However, the game-related factors such as payment condition, goodness of the suggestion and reputation played a role in defining the condition under which such a behavioral rule is more or less used. In both game settings (Payment On/Off), we observed how a negative reputation determined a lower use of the reputational criterion. In other words, in their feedback decision making, participants appeared to rely more on the reputation of their partner when the Observer's reputational score was positive compared to when it was negative. In the same way, and in both game settings, bad suggestions seemed to undermine the adoption of the reputation criterion to feedback, while in the case of bad suggestions provided by negatively rated Observers, we registered an increase in such use. Interestingly, the Receivers did not use the reputation criterion differently, only in relation to the type of reputation, as we could see from the insert in Figure 1. Indeed, reputation revealed complex relationships both with the Payment and the goodness of the suggestion and thus affected the use of the reputation criterion through the interaction effects (confirming H2).

3.4 | Reputation inertia as deviation from rationality

The type of suggestion received (i.e., good or bad) influenced the feedback tendency to rely on the reputational information in both game settings (Figure 1). However, the probability of providing feedback in line with the previous reputation of the Observer was higher in those cases in which the goodness of the suggestion and the reputation score were concordant. Instead, when these two observables were discordant (i.e., positive reputation-bad suggestion, negative

TABLE 6 Inertia table

Experimental inertia			
Condition	Payment off	Payment on	Rational agents
Reputation(−)[a] coherence(−)	0.83	0.91	1
Reputation(+)[a] coherence(−)	0.57	0.47	0
Reputation(−)[a] coherence(+)	0.61	0.45	0
Reputation(+)[a] coherence(+)	0.88	0.83	1

Note: In the table the average values of the variable inertia where $I \in (0, 1)$ in both game settings and the probability of using the reputational criterion that would be expected from a rational agent are reported.
[a]Underlines the interaction between the variables.

reputation-good suggestion), the probability of adhering to the reputational criterion was lower.

In the first cases, the use of information based on reputation was not different (and diversifiable from a behavioral point of view) from what a rational agent would have done (i.e., give feedback based on the behavior of the interactor). Instead, in the latter cases in which the observation goodness and the reputation scores were discordant, feedback that maintained and confirmed the reputation of the Observer breached the principle of rationality. To better represent this irrational component that we could call properly "reputation inertia" we subtracted the probability of feedback that we would expect from a rational agent $p_I(R)$ from the probability of feedback based on reputation registered within our sample $p_I(S)$ (Table 6).

Since in our game the participants were asked to provide a feedback only after becoming aware of the actual behavior of their interactor (i.e., when the participants had all the information about the specific event they have to evaluate), we considered "irrational" those participants which relied on the reputation of their interactor when reputation score and behavior were discordant (e.g., the Receiver provided a positive feedback to an Observer with a positive reputation immediately after having received a damage from him).

FIGURE 2 In the figure is reported the dynamics of "irrational inertia," defined as those events in which the Receiver rated the Observer just following her/his reputation, but against the "rational" evaluation that should be derived from the real behavior of the Observer itself. Again, in red and black are reported the average values and standard error of the mean, and the percentage of "irrational inertia" is reported separately for the four possible interaction typologies. In the insert figure, the average "irrational inertia" is reported with respect to the goodness of the Observer's suggestion
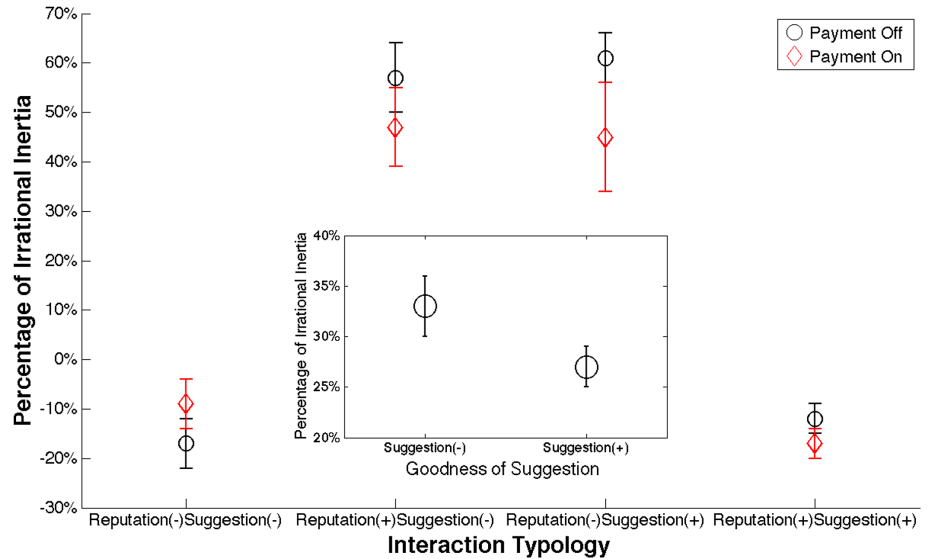

TABLE 7 Best generalized linear mixed model (GLMM) predicting the "irrational inertia" behavior

GLMM best model "irrational inertia"				
	Model precision	Akaike*	F	Df-1(2)
Best model	75.7%	47.763	34.16***	6 (565)
Fixed effects				
Factor		F		Df-1(2)
Payment[a] reputation		5.28*		1 (565)
Payment[a] goodness of suggestion		2.91*		1 (565)
Reputation[a] goodness of suggestion		153.01***		1 (565)
Parameter		Coefficient (β)		Student t
Payment(0)[a] reputation(−)		2.078		9.88***
Payment(1)[a] reputation(−)		1.585		5.36***
Payment(0)[a] goodness of suggestion(−)		2.168		8.54***
Payment(1)[a] goodness of suggestion(−)		1.613		6.51***
Reputation(−)[a] goodness of suggestion(−)		−4.122		−12.37***

Note: (0): Payment not present.
[a]Underlines the interaction between the variables.
*$p < .05$.; ***$p < .001$.

TABLE 8 Best generalized linear mixed model (GLMM) predicting the "irrational inertia" behavior with respect to the "typology of interaction"

Best GLMM				
	Model precision	Akaike*	F	Df-1(2)
Best model	75.7%	27.948	33.12***	3 (568)
Fixed effects				
Factor		F		Df-1(2)
Typology of interaction		80.74***		1 (568)
Payment[a] typology of interaction		2.61*		2 (568)
Parameter		Coefficient (β)		Student t
Payment(0)[a] concordant		−1.730		−5.15***
Payment(0)[a] discordant		0.529		1.94*
Payment(1)[a] discordant		−1.585		−4.88***

Note: The typologies of interaction can be concordant if the action of the Observer is congruent with her reputation (e.g., a fair advice from an Observer with good reputation), and discordant otherwise. (0): Payment not present.
[a]Underlines the interaction between the variables.
*$p < .05$.; ***$p < .001$.

3.4.1 | Irrational feedback model

The subtraction ($p_I(S) − p_I(R)$) allowed us to represent the irrationality for all the cases considered. Indeed, two types of irrationality were derived. For the concordant cases, the irrationality manifested itself when the Receiver did not follow reputation criterion when he should have done it. Thus, for instance, the Receiver gave a dislike to an Observer with a good reputation that helped him. This type of irrational feedback does not contribute in any way to the "reputation inertia" and actually hurt the maintenance of the Receiver's reputation. For the discordant cases, the irrationality reflected, as already pointed out, the reputation inertia phenomenon, which is considered by H3.

Figure 2 shows the degree of deviation from rationality due to the Irrational Inertia in every condition and for both game settings.

In Table 7, the best model for the Receiver's irrationality is presented.

Three interaction effects involving reputation, goodness of suggestion, and payment determined the levels of Irrational Inertia within
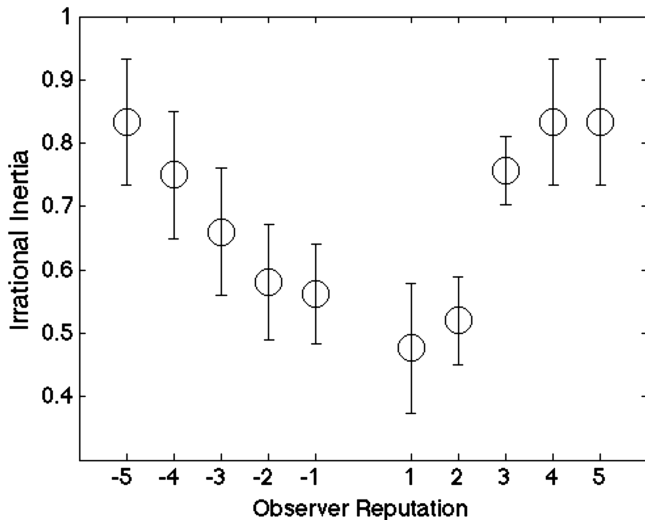
FIGURE 3 In the figure the irrational inertia (II) is reported as a probability (i.e., a normalized frequency of occurrence) in relation with the Observer reputation. The magnitude of the II appears as clearly related with the magnitude of the Observer's reputation

our game. As expected, the payment condition reduced the levels of irrationality both in relation to a negative reputation and to a bad suggestion. In general, when disjointed, these two factors seemed to increase the Irrational Inertia frequency as well as (as seen already in Table 5) the use of the reputational criterion (see also the insert in Figure 2). Hence, the increase in irrationality due to these factors is attributable to those cases in which the Receivers did not consider either reputation or personal experience (i.e., concordant cases irrationality). However, when a negative reputation and a bad suggestion were joined, this combination of events led to a decrease in the feedback irrationality.

3.4.2 | Irrationality among the different typologies of interaction

To better present the same phenomenon from another point of view (i.e., concordant vs. discordant cases), we produced a new model (Table 8) involving the Typology of Interaction as a parameter.

We did not consider goodness of the suggestion and reputation in this new model since the variable Interaction is a linear composition of them. We observed how the degree of irrationality was lower in the concordant cases when the game setting did not involve a cost to make a suggestion. Moreover, the payment reduced the irrationality in the discordant cases and thus hindered the irrational reputation inertia (H4). However, this cost also increased the irrationality in the concordant cases (and specifically for a positive reputation and good suggestion combination as we can appreciate from Figure 2).

Moreover, we analyzed the relationship between the reputation values achieved by the Observer and the degree of irrational inertia shown by the Receivers (Figure 3).

Observers' experimental reputation values ranged from −10 to +10. These values were matched in couples (e.g., −10 and −9 values defined

the −5 reputation level) to obtain five levels for each typology of reputation (i.e., positive and negative). We registered a non-linear relationship between the Observers' reputation level and the tendency to feedback according to reputation in the discordant cases. Such non-linear relation can be approximated to a quadratic or a complex function. In general, we observed the existence of two plateaux in correspondence to the reputation limit values (i.e., −5 and +5) and of a minimum point for those levels proximate to zero (i.e., −1 and +1). Notably, our participants gave a bad evaluation to a very badly rated Observer who provided a good suggestion nearly the 85% of the time. The same level of irrationality happened with a very good rated Observer who gave a bad hint to the Receiver. In both of these cases, the high level of reputation triggered irrational feedback from the Receiver.

4 | DISCUSSION

Overall, our work contributes to clarifying which mechanisms are involved in reputation building and maintenance. We showed how reputation is really "made" within a widespread feedback system (e.g., e-commerce sites). As predicted by H1, individuals did not behave like rational agents. In their feedback, they considered not only the direct behavior of their social partner (i.e., the goodness of his/her suggestion) but also they were significantly influenced by the previously acquired reputation of their interactor. In line with the work of Semmann et al. (2005), we discovered how reputation could exert a social influence by also directing social rewards (i.e., the reputation itself) in a manner disconnected from personal experience. A good reputation attracted other positive feedback, and surprisingly, this happened even when the Observer's behavior damaged (bad suggestion) the Receiver. The opposite also appeared true. A negative reputation attracted other negative feedback more frequently, and this occurred even when the "bad" Observer provided a good suggestion. The fact that this effect persisted even when the observation goodness and the reputation score were discordant makes reputation resistant to change. Therefore, our results supported H2 and H3.

Furthermore, the tendency to use the reputational information as one criterion for the feedback action resulted independent from sociodemographic and psychological factors, while the game scenario (i.e., Payment On/Off) influenced its use. When the information transmission (i.e., the Observer suggestion) was free, the tendency to feedback according to the previous reputation was not different for positive and negative reputations. Whereas, if the passage of information from the Observer to the Receiver entailed a cost, we observed a new way of using the reputation to adjust feedback. In this latter game scenario, Receivers seemed to be more influenced by a positive reputation in their feedback action and less by a negative one. In other words, in this condition, a good-rated Observer seemed more able to exert an influence on the Receivers' feedback behavior, whereas badly-rated Observers were less likely to be evaluated by Receivers using the reputation rule (i.e., give a dislike disregarding personal experience). However, the analysis of the irrational component of the reputation inertia phenomenon (i.e., the proper reputation inertia) showed that there was no difference

between the positive and the negative reputation limit values. In the discordant cases, higher reputation values determined a higher probability of engaging in irrational feedback. This probability decreased for lower levels of reputation.

Moreover, the payment allowed real behavior to be considered more within the context of the Receivers' decision making. In other words, as predicted by the costly signaling theory (Barclay, 2011), the cost involved to provide data activated in the Receivers a more data-driven (i.e., more focus on the social partner's behavior) cognitive evaluation process. Indeed, willingness to pay provides useful information to others both when the interactor helped (Albert et al., 2007) and when he damaged others (Herrmann, Thöni, & Gächter, 2008). This additional data-driven information contributed to determining the reputation of the Observers. Indeed, the payment action seemed to establish a new pattern of reputation influence, conferring importance to personal experiences as well.

We analyzed the impact of the payment on those cases in which the observation goodness and the reputation score were discordant (i.e., those that maintained the reputation against the real behavior). We observed a reduction in the usage of the reputation criterion by our participants. In a sense, the payment made individuals more rational in their general feedback. However, even if reduced in frequency, the reputation inertia persisted. Such a result fits with some novel findings of reputation side-effects (Frey & Van De Rijt, 2016). As pointed out by the two authors, reputation exhibits a sort of cumulative advantage that is disconnected from the trustworthiness of the interactors, and this simply led to strongly preferring the individuals with good reputation. Similarly, we provided evidence of a similar cumulative effect (i.e., reputation inertia effect) in the reputation building. In our game, the interactions were scheduled and were not based on personal preferences, but still, individuals frequently chose to build the reputation of their social partner according to their previously acquired reputation and not in response to the actual behavior, even if this conduct ended up damaging the Receiver.

To conclude, people appear to rely significantly on the previously acquired reputation once they have to "rate" a partner during on-line social interactions. i.e., the probability of receiving a like within an online reputation system appears as positively (even if not linearly) related with the previous amount of likes obtained by the target (e.g., her/his reputation). Such a tendency (or heuristics) seems to be very pervasive in virtual environments, probably because of the lack of information affecting the decision making processing (Friedland, 1990), and produces a bias that we labeled as "reputation inertia." The absence of any effect referable to age, gender and psychological variables can be explained by the Social Identity Model of Deindividuation Effects (Postmes & Spears, 1998). Indeed, as emerged in a recent work, when people experience the psychological state defined as "deindividuation" in response to anonymous virtual group interaction, they seem to be more influenced by reputation than by their individual characteristics (Duradoni, Paolucci, Bagnoli, & Guazzini, 2018).

Since "reputation inertia" appears to be able to distort the reputation dynamics of a web-based social system, so reducing its effectiveness and robustness, the modeling of the virtual human dynamics underlying this phenomenon could proceed to exploit the preliminary findings of this paper. Indeed, virtual environments that can take into account the impact of the reputation on individuals' social judgment could foster citizens public reason and social coordination capabilities (Condor, 2011) as well as benefiting all those projects that use reputation systems to cope with free-riding and social loafing dynamics (e.g., Collective Awareness Platforms, crowdsourcing projects).

ORCID

Mirko Duradoni https://orcid.org/0000-0001-8272-9484

REFERENCES

Albert, M., Güth, W., Kirchler, E., & Maciejovsky, B. (2007). Are we nice (r) to nice (r) people?—An experimental analysis. *Experimental Economics*, *10*(1), 53–69. https://doi.org/10.1007/s10683-006-9131-3

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, *27*(5), 325–344. https://doi.org/10.1016/j.evolhumbehav.2006.01.003

Barclay, P. (2011). The evolution of charitable behaviour and the power of reputation. In C. Roberts (Ed.), *Applied evolutionary psychology* (pp. 149–172). Oxford, England: Oxford University Press ISBN: 978-0-19-958607-3.

Bidgoly, A. J., & Ladani, B. T. (2016). Benchmarking reputation systems: A quantitative verification approach. *Computers in Human Behavior*, *57*, 274–291. https://doi.org/10.1016/j.chb.2015.12.024

Biel, A., & Thøgersen, J. (2007). Activation of social norms in social dilemmas: A review of the evidence and reflections on the implications for environmental behaviour. *Journal of Economic Psychology*, *28*(1), 93–112. https://doi.org/10.1016/j.joep.2006.03.003

Chiregi, M., & Navimipour, N. J. (2016). A new method for trust and reputation evaluation in the cloud environments using the recommendations of opinion leaders' entities and removing the effect of troll entities. *Computers in Human Behavior*, *60*, 280–292. https://doi.org/10.1016/j.chb.2016.02.029

Condor, S. (2011). Towards a social psychology of citizenship? Introduction to the special issue. *Journal of Community & Applied Social Psychology*, *21*(3), 193–201. https://doi.org/10.1002/casp.1089

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, *49*(10), 1407–1424. https://doi.org/10.1287/mnsc.49.10.1407.17308

Dolle, R. (2014). *Online reputation management* (BS thesis). Faculty of Management and Governance.

Duradoni, M., Paolucci, M., Bagnoli, F., & Guazzini, A. (2018). Fairness and trust in virtual environments: The effects of reputation. *Future Internet*, *10*(6), 50. https://doi.org/10.3390/fi10060050

Frey, V., & Van De Rijt, A. (2016). Arbitrary inequality in reputation systems. *Scientific Reports*, *6*(38304). https://doi.org/10.1038/srep38304

Friedland, N. (1990). Attribution of control as a determinant of cooperation in exchange interactions. *Journal of Applied Social Psychology*, *20*(4), 303–320. https://doi.org/10.1111/j.1559-1816.1990.tb00413.x

Giannini, M., Pannocchia, L., Grotto, L., & Gori, A. (2012). A measure for counseling: The five factor adjective short test (5-fast). *Counseling Giornale Italiano di Ricerca e Applicazioni*, *3*, 384.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367. https://doi.org/10.1126/science.1153808

Hutton, J. G., Goodman, M. B., Alexander, J. B., & Genest, C. M. (2001). Reputation management: The new face of corporate public relations? *Public Relations Review*, *27*(3), 247–261. https://doi.org/10.1016/S0363-8111(01)00085-6

Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal processes. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action* (pp. 195–213). Washington, DC: Hemisphere Publishing Corp.. ISBN: 978-1-56032-269-6.

McCulloch, C. E., & Neuhaus, J. M. (2001). Generalized linear mixed models. *Encyclopedia of Biostatistics*. https://doi.org/10.1002/0470011815.b2a10021

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring/the dynamics of indirect reciprocity. *Nature*, 393(6685), 573–577. https://doi.org/10.1038/31225

Ohtsuki, H., & Iwasa, Y. (2004). How should we define goodness?— Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology*, 231(1), 107–120. https://doi.org/10.1016/j.jtbi.2004.06.005

Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, 123(3), 238–259. https://doi.org/10.1037/0033-2909.123.3.238

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature*, 5, 3677. https://doi.org/10.1038/ncomms4677

Rovai, A. P., Wighting, M. J., & Lucking, R. (2004). The classroom and school community inventory: Development, refinement, and validation of a self-report measure for educational research. *The Internet and Higher Education*, 7(4), 263–280. https://doi.org/10.1016/j.iheduc.2004.09.001

Schneider, J., & Cook, K. (1995). Status inconsistency and gender: Combining revisited. *Small Group Research*, 26(3), 372–399. https://doi.org/10.1177/1046496495263004

Semmann, D., Krambeck, H.-J., & Milinski, M. (2005). Reputation is valuable within and outside one's own social group. *Behavioral Ecology and Sociobiology*, 57(6), 611–616. https://doi.org/10.1007/s00265-004-0885-3

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(44), 17435–17440. https://doi.org/10.1073/pnas.0704598104

Van Der Heide, B., Johnson, B. K., & Vang, M. H. (2013). The effects of product photographs and reputation systems on consumer behavior and product cost on ebay. *Computers in Human Behavior*, 29(3), 570–576. https://doi.org/10.1016/j.chb.2012.11.002

Wæraas, A., & Byrkjeflot, H. (2012). Public sector organizations and reputation management: Five problems. *International Public Management Journal*, 15(2), 186–206. https://doi.org/10.1080/10967494.2012.702590

AUTHOR BIOGRAPHIES

Mirko Duradoni is a PhD student at the Information Engineering Department and a member of the Virtual Human Dynamics Laboratory at the University of Florence, Italy. He holds a master's degree in psychology from the Department of Education and Psychology, University of Florence, Italy. His main research interests concern the role of social norms and reputation in influencing online behaviors and the psychosocial ergonomics of web-based systems.

Giorgio Gronchi is an assistant professor at University of Florence. He got a PhD in Psychology in 2009 from University of Florence. He currently works in the Cognitive Psychophysiology Lab of the Department of Neuroscience, Psychology, Drug Research and Child Health (University of Florence). His research interests include probabilistic reasoning, randomness perception, decision making as well as mathematical and computational models of cognition. He has worked as a professional psychologist in the field of human–computer interaction, web-usability and applied cognitive psychology.

Leonardo Bocchi received the PhD degree in Biomedical Engineering from the University of Bologna, Italy. He is Associate Professor in Biomedical Engineering with the Department of Information Engineering, University of Firenze, Firenze, Italy. His research interests include Computer Aided Diagnosis tools based on medical image and signal processing, artificial intelligence techniques based on soft computing, and modelling of physiologic systems and cognitive processes. Main applications concern the detection and classification of breast tumors, based on ultrasound, radiographic or magnetic resonance imaging; the study of the micro circulatory dynamics and their relations with chronical or complex pathological conditions as diabetes or sepsis; and cognitive and behavioral modelling using nonintrusive monitoring techniques as movement analysis, face expression properties, and sentiment analysis. He is member of the IEEE-Engineering in Medicine and Biology Society and member of the IEEE-EMBS Technical Committee on Cardiopulmonary Systems. He is author and coauthor of about 100 papers in peer-reviewed journals and proceedings of national and international conferences.

Andrea Guazzini received his PhD in Complex system and nonlinear dynamics by University of Florence). From 2008 to 2011 he was postdoc at the Institute of Informatics and Telematics of the National Research Council of Italy. In 2012 he became researcher at the department of Education and Psychology, University of Florence, and responsible for the laboratory for the study of the human virtual dynamics (VirtHuLab) at the Centre for the Study of Complex Dynamics of the University of Florence. Finally he was involved in several EU projects, exploiting the VirtHuLab to get experiments and laboratory research, and responsible for EU project Horizon 2020

"PROVA" regarding the radicalization virtual dynamics. His research interests are in the fields of the experimental and cognitive psychology, neuropsychology, social cognition, with a particular interest in the virtual social dynamics. He worked on the connection between the statistical mechanics and the nonlinear physics to cognitive and social psychology. He is the author of about 90 contributions in international refereed journals.