Fake online reviews:

a study on eWOM influence when suspicions arise

Nicoletta Cavazza

Margherita Guidetti

Università di Modena e Reggio Emilia

Corresponding author:

Nicoletta Cavazza, Università di Modena-Reggio Emilia, via Allegri 9, 42121 Reggio Emilia, Italy.

Phone: +39 0522 523262; fax: +39 0522 523055; e-mail: nicoletta.cavazza@unimore.it

We wish to express our gratitude to Carlotta Masselli for helping us collecting the data.

Abstract

Online reviews are widespread and can strongly affect consumer choice. However, the audience may know that these tools can be used and counterfeited for propaganda goals. We present an experiment 2 (falsity suspicion vs. control condition) x 2 (valence order: positive reviews first vs. negative reviews first) factorial design aimed at exploring what happens when people get suspicious about reviews' authenticity. As expected, results showed that suspicion hinders careful information processing. In addition, it affected restaurant evaluation by increasing the number of positive reviews hypothesized as fake, which in turn reduced positive reviews perceived as useful,

dragging the judgement toward the negative pole (partial mediation). The implications are discussed.

Keywords: Social influence, eWOM, suspicion, consumer, online review.

Fake online reviews: a study on eWOM influence when suspicions arise

Word of mouth (WOM) is a powerful tool of social influence: indeed, in the consumer choice domain, it is known that people take into great consideration the evaluations provided by acquaintances who already tried a certain product (e.g., Richins, 1983). Research shows that WOM entails long term effects, and that individuals perceive the source as believable and reliable (Bone, 1995; Herr, Kardes, & Kim, 1991).

With the advent of the Internet, comments and reviews about products in general, and restaurants in particular, have become widespread through dedicated websites. This phenomenon is encapsulated in the expression electronic WOM (eWOM). Thanks to these tools, consumers have at hand a great amount of information for orienting their choices, reducing their uncertainty, risks of delusions and costs at any time (Chatterjee, 2001; Hu, Liu, & Zhang, 2008; Park & Lee, 2008).

The implementation of dedicated websites for consumer reviews reinforces the effect of traditional WOM because these platforms enormously broaden the circle of the providers of information well beyond personal networks of friends, acquaintances and colleagues. In addition, eWOM is provided in written form and thus is suitable for a more close examination compared to the oral form of WOM (Lee, Park, & Han, 2008). Indeed, research has shown that eWOM effects are even more powerful than advertising (*inter alia*, Bickart & Schindler, 2001; Smith, Menon, & Sivakumar, 2005; Trusov, Bucklin, & Pauwels, 2009).

However, a critical and neglected difference between WOM and eWOM is that when a relative or an acquaintance tell us s/he is very satisfied by a product, we do not suspect that s/he has been paid by the producer for spreading positive information about its merchandise. On the contrary, the idea that online reviews might be falsified in order to influence costumers' choice is very popular, insomuch as algorithms distinguishing fake from authentic reviews have been developed (Mukherjee, Liu, & Glance, 2012; Ott, Choi, Cardie, & Hancock, 2011).

We conducted a study aimed at investigating what happens when people read online reviews (of a restaurant in our case) under falsity suspicion. In other words, we explored whether suspecting that a review is fake would affect the degree of information processing and the final evaluation of the reviewed product.

The impact of falsity suspicion

Newspapers have often published news or reports about the presence of fake online reviews of hotels, restaurants and travelling services (e.g., Vincenzi, 2012, September 2). Thus, the general public knows that these tools can be used and counterfeited to propagandise goods and services.

There are two kinds of authors of fake reviews. On the one hand, there are individuals motivated by a personal interest (typically the owners) in promoting or discrediting (as unfair competition or revenge) a product or a service by writing falsified evaluations. On the other hand, the increasing use and the proved efficacy of eWOM has fostered the development of a market of reviews: nowadays a commercial enterprise can contact promotion agencies that, among other services, offer to improve the reputation of a product or a service by inserting a massive amount of focused and monitored information and judgments. In both cases, counterfeit reviews can be either positive or negative, thus the comments' valence is not a good clue to estimate the authenticity of the information.

The awareness that not all the reviews are necessarily genuine induces the user to face two evaluation tasks: a) inferring the features and qualities of the product to be selected, and b) trying to figure out the source's intentions, that is ascertain whether the available information is reliable or not (Racherla, Mandviwalla, & Connolly, 2012).

What happens then when the readers suspect that some reviews might be fake? To the best of our knowledge, there are no studies directly asking this question. However, in order to guide our hypotheses, we collected suggestions from the research about the potential influence of alleged or admitted false information. Some studies have shown that even explicitly declared deceitful information can influence people's judgements when they rely on automatic cognitive processes

(Dimofte & Yalch, 2005). Indeed, time scarcity or cognitive load drives individuals to not weight true and false information differently in decisional processes (Gilbert, Tafarodi, & Malone, 1993). For instance, people tend to laugh automatically when watching TV shows including canned laughter (Chapman, 1973), but the effect weakens if the laughter is recognised as artificial (Lawson, Downing, & Cetola, 1998).

In the advertisement domain, it has been shown that people exposed to an ad claim declared as false can remember it as true after three days (Skurnik, Yoon, Park, & Schwarz, 2005). However, an alert inducing misinformation suspicion just prior to the exposure of a message elicited resistance to persuasion through a slow and careful reading of the information, whereas the same alert had no effect when placed after the message (Greene, Flynn, & Loftus, 1982).

Beyond the influence exerted by the dubious message content, knowing or suspecting that a review is counterfeit leads the reader to deduce the source's persuasive intention. Many studies in the persuasion domain have shown that this awareness plays a critical role in the process outcome by activating reactance and motivating resistance to the persuasion attempt (e.g., Campbell & Kirmani, 2000; Fein, Hilton, & Miller, 1990; Friestad & Wright, 1994; Hovland, Janis, & Kelley, 1953; Walster & Festinger, 1962). Furthermore, some studies analysed how the knowledge of a source's manipulative purposes affects the depth of message elaboration. In particular, Echebarria-Echabe (2010) recently found that receivers tend to reduce the cognitive effort of message processing when they suspect the propagandistic goal of the source.

To summarise, we can affirm in general that false information affects audience's judgments and behavioural intentions only when they are formulated through automatic processes. In addition, inferring the source's persuasive intention can engender resistance and an early closure of the message processing.

Finally, we need to consider that, as anticipated above, although fake reviews can be either positive or negative, these different kinds of information do not have the same impact on people's evaluations: indeed, negative information is generally perceived as more diagnostic and useful for

judgments formation (Rozin & Royzman, 2001). In line with this "negativity bias", negative reviews are considered more believable than positive ones, thus they have a greater impact on costumers' choices, although this was found to be true for instrumental products only and not for hedonic products (Sen & Lerman, 2007).

In addition, and more generally, a primacy effect emerged in previous studies: regardless of number and valence of available information, the valence of the first review influenced the final judgment more than the subsequent ones (Coker, 2012). However, given that the primacy effect emerged particularly in shallow processing conditions (Webster, Richter, & Kruglanski, 1996), we can expect that the role of positive and negative reviews would be more balanced when people examine them deeply.

Overall, research so far has not clarified whether the widespread doubts about the authenticity of online reviews lead the reader to less carefully process their content, whether this suspicion concerns positive reviews in particular, as they are generally perceived as less believable, whether this undermines their perceived usefulness for judgement formation, and finally whether the process elicited by suspicion extends to the final evaluation.

Overview and hypotheses

In order to answer these questions, we conducted an experimental study with a 2 (falsity suspicion vs. control condition) x 2 (information valence order: positive reviews first vs. negative reviews first) factorial design. Participants read six reviews about a bogus restaurant, called «La prima declinazione», located in the tourist town Pietrasanta (far from the data collection place). The reviews were selected through a pre-test (see below) to include three positive and three negative reviews. Before participants read the comments, half of them had been warned that some reviews could be fake. The order of reviews was counterbalanced so that half of the sample read the three positive comments first and the remaining half read the negative comments first.

Drawing on the reviewed literature above, we expect that:

Hp1. Falsity suspicion would impede a careful processing of the reviews' content and thus would elicit a primacy effect on the final judgement.

Hp2. Falsity suspicion would affect restaurant evaluation, mediated by the number of positive reviews detected as fake and their perceived utility, in sequence. In particular, we will test a mediational model where falsity suspicion drives detection of more fake positive reviews (than in the control condition) and this worsens the appraisal of their usefulness, that instead would be positively associated with the final judgment about the restaurant. By virtue of the negativity bias which makes people consider negative information more believable and diagnostic than positive, we might expect that negative reviews do not yield the same influential process on the restaurant's final judgment. However, since previous studies suggested that the negative bias only emerges for instrumental goods and a restaurant is a hedonic service, we could find the same process for negative reviews as well. This is why we will test the two models separately.

Method

Participants

Forty-six women and 34 men, aged 25 to 89 years old (M = 47.77; SD = 13.53) took part in our study. They were recruited through personal mailing lists and snowball sampling by selecting adults that usually go to restaurants (at least once a month). Participants were randomly assigned to one of the four experimental conditions in our 2 (alert about the potential presence of fake reviews vs. no alerts) x 2 (information valence order: positive reviews first vs. negative reviews first) factorial design.

Procedure

Participants were emailed with the request of completing an attached questionnaire. On the first page, they were asked to imagine they were looking for information about a restaurant named «La prima declinazione» in Pietrasanta and they found, in a dedicated website, six reviews of customers who had already gone to the restaurant. In the experimental condition, this first page also included

the sentence: «Imagine you have heard that the reviews published on that website can be fake» for prompting suspicion, whereas that warning did not appear in the control condition.

Before reading the reviews, participants were asked to report, on a 5-point scale where 1 = *not at all* and 5 = *very much*, how useful they generally considered online reviews for the choice of a restaurant. Then they were asked to carefully read the six reviews and answer a series of questions operationalizing the examined constructs (see below) and demographic data.

Drawing on various restaurant review dedicated websites, we composed three overall positive and three negative reviews. In order to assure that the positive and negative valence would be correctly perceived and that all the reviews would be evaluated as equally realistic, we pre-tested them with 20 participants. Actually, the analysis of variance confirmed that the positive reviews were associated with a more positive judgment of the restaurant than the negative ones, $F(1, 18) = 276.11$, $p < .001$. On the contrary, the repeated measures analysis of variance on realism perception elicited the expected non-significant result, $F(5, 95) = .68$, $p = .64$.

Besides manipulating the falsity suspicion, we manipulated the reviews' valence order as well, in order to identify the possible primacy effect, due to the reluctance to carefully process the information. Following Coker's (2012) procedure, participants read the three positive reviews first in one condition and the three negative reviews first in the other.

Measures

Reviews' falsity judgement. For each review, we asked participants whether they believe it was either true or fake. Based on these answers, we computed the number of positive and negative reviews considered as fake.

Reviews' perceived usefulness. For each review, we asked participants how useful it was in deciding whether to go to that restaurant. Based on these answers, we computed a usefulness index for both negative ($\alpha = .63$) and positive ($\alpha = .73$) reviews.

Global evaluation of the restaurant. This index was computed as the mean of three items: the first asked participants how many «stars» from 1 to 5 they would give to the restaurants, based

on the information they read; the other two items assessed the desire and the likelihood of dining in that restaurant in case they went on holiday to Pietrasanta ($\alpha = .87$).

Results

Manipulation Check

On the whole, participants detected a mean of 1.52 fake reviews ($SD = 1.10$), but this estimate was significantly higher in the falsity suspicion condition ($M = 2.02$, $SD = 1.29$) than in the control condition ($M = 1.02$, $SD = .53$), $F(1,78) = 20.54$, $p = .001$, $\eta^2_p = .21$.

The influence of suspicion on restaurant evaluation

In order to test the first hypothesis, we ran as analysis of variance (ANOVA) with two independent factors (falsity suspicion and reviews valence order) on the global evaluation of the restaurant. Falsity suspicion main effect was significant, $F(1,76) = 41.69$, $p < .001$, $\eta^2_p = .35$, as well as the interaction effect, $F(1,76) = 9.01$, $p = .004$, $\eta^2_p = .11$.[1] Overall, falsity suspicion worsened the restaurant evaluation compared to the control condition. In addition, as expected, information valence order influenced the judgement in the suspicion condition, $F(1,38) = 7.05$, $p = .012$, $\eta^2_p = .16$, but not in the control condition, $F(1,38) = 2.12$, $p = .15$, $\eta^2_p = .05$.

Looking at the means of the restaurant evaluation in the different conditions (Table 1), we can confirm that the inducement of doubts and distrust yielded a significant primacy effect: indeed, only in this condition did participants form their judgments with a strong anchoring to the first reviews they read: They expressed a more positive judgement when positive reviews were presented first, and a more negative judgement when negative reviews were presented first.

Mediational analysis

In order to understand how falsity suspicion worsens restaurant evaluation, we performed the two mediational models elucidated in the second hypothesis: one focused on the number of positive reviews detected as fake and their perceived usefulness, the other on the same features of negative reviews. To this end, we used PROCESS, the SPSS macro provided by Hayes (2012), running

model 6 with 5000 bootstrap resamples. This analysis allows testing three indirect paths: the first includes only the number of fake reviews as a mediator, the second includes only their perceived usefulness as a mediator, and the third includes the complete sequence depicted in Figure 1.

In the first model, the manipulation of suspicion (compared to the control condition) increased the number of positive reviews detected as fake (b = .72, SE =.17, t = 4.13, p < .001), which in turn reduced the perceived usefulness of positive reviews (b = -.42, SE = 13, t = -3.29, p = .001) that consequently worsened the global evaluation of the restaurant (b = .36, SE = .10, t = 3.48, p < .001). The analysis of the indirect effects confirmed that the only significant model was the one including the complete sequence (indirect effect = -.11; SE = .05; LL =-.27; UL = -.03; Figure 1). It is, however, a partial mediation, as the direct effect of suspicion on restaurant evaluation remained significant while controlling for the two mediator effects (b = -1.00, SE = .20, t = -5.05, p < .001).

We tested the same model including the mediators concerning negative reviews (number of negative reviews detected as fake and their perceived usefulness). No significant indirect effect emerged in this case.

Discussion

So what happens when a user of online reviews suspects they are fake? According to our results, the first impact was exerted on information processing depth: our suspicious participants expressed a restaurant judgment anchored to the valence of the first reviews read suggesting that they quickly closed their information analysis. In addition, our respondents focused their doubts on positive reviews, considering them more likely to be fake than negative ones, and then less useful for judgement formulation. As expected, the falsity suspicion affected restaurant evaluation by increasing the number of positive reviews detected as fake, which in turn reduced the usefulness of positive reviews and thus gave way to negative reviews dragging the judgement toward the negative pole.

We cannot say why the mere warning about the possible presence of fake reviews was only in part mediated by this process and in part directly worsened their final judgment. As Fein and Hilton (1994) suggested, this direct effect might be due to the negative impression (engendered by falsity suspicion) about the restaurant owner, which would lead to a negative evaluation of the service s/he provides. Unfortunately, the present data did not allow us to test this interpretation. Therefore, further studies should include this variable as well. In addition, the current study did not allow participants to choose which reviews they wanted to read, thus it would be interesting to let participants have free to access to single reviews from a title list. This procedure would allow to better understand how the information processing develops in the various conditions.

Notwithstanding these limitations, our study contributes to clarifying that suspicious, in addition to directly affecting the final judgment, elicits scepticism that hinders the careful scrutiny of the presented information. This obstacle does not equally concern all the available information, but paves the way for the evaluation process by weakening the role of positive information while maintaining the strong impact of negative ones.

The understanding of this process also has practical implications: indeed, our study suggests that strategies aimed at improving one's own business reputation cannot be simply based on the insertion of positive reviews and comments for two reasons. First, the potential immediate advantage of fake positive reviews turns into long term damage for the entire evaluation system that loses credibility when falsity suspicion arises. Furthermore, increasing positive information may be totally useless when facing a sceptical recipient who is unwilling to carefully process them. Therefore, the position of information with different valences might work better than the mere proportion of positive and negative reviews.

References

Bickart, B, & Schindler, R. M. (2001). Internet Forums as Influential Sources of Consumer Information. *Journal of Interactive Marketing, 15,* 31-40.

Bone, P.F. (1995). Word-of-mouth effects on short-term and long-term product judgments. *Journal of Business Research, 32,* 213-223.

Campbell, M. C., & Kirmani, A. (2000). Consumers' use of persuasion knowledge: The effects of accessibility and cognitive capacity on perceptions of an influence agent. *Journal of Consumer Research, 27,* 69-83.

Chapman A. (1973). Funniness, jokes, canned laughter and recall performance. *Sociometry, 36,* 569-578.

Chatterjee, P. (2001). Online reviews: do consumers use them? *Advances in Consumer Research, 28,* 129-133.

Coker, B. L. (2012). Seeking the opinions of others online: Evidence of evaluation overshoot. *Journal of Economic Psychology, 33,* 1033–1042.

Dimofte, C.V., Yalch, R.F. (2005). Consumer responses to false information: Is believability necessary for persuasion? In F. R. Kardes, Herr, P. M., & J. Nantel (Eds.) *Applying social cognition to consumer-focused strategy* (pp. 281-296). Mahwah, NJ: Lawrence Erlbaum Associates.

Echebarria-Echabe, A. (2010). Effects of suspicion on willingness to engage in systematic processing of persuasive arguments. *The Journal of Social Psychology, 150*, 148-159.

Fein, S., & Hilton, J. L. (1994). Judging others in the shadow of suspicion. *Motivation and Emotion, 18,* 167-198.

Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motivation and the correspondence bias. *Journal of Personality and Social Psychology, 58,* 753-764.

Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of consumer research, 21,* 1-31.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of personality and social psychology, 65,* 221-233.

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior, 21,* 207-219.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modelling [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research, 17,* 454-462.

Hilton, James; Fein, Steven; Miller, Dale (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin 5,* 501-512.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion; psychological studies of opinion change.* New Haven, CT: Yale University Press.

Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management, 9,* 201-214.

Lawson, T. J., Downing, B., & Cetola, H. (1998). An attributional explanation for the effect of audience laughter on perceived funniness. *Basic and Applied Social Psychology, 20,* 243-249.

Lee, J., Park, D. H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications, 7,* 341-352.

Mukherjee, A., Liu, B., & Glance, N. (2012, April). Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web* (pp. 191-200). ACM.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 1,* 309-319.

Park, D.H., & Lee, J. (2008). EWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications, 7,* 386–398.

Racherla, P., Mandviwalla, M., & Connolly, D. J. (2012). Factors affecting consumers' trust in online product reviews. *Journal of Consumer Behaviour, 11,* 94-104.

Richins, M. L. (1983). Negative word-of-mouth by dissatisfied consumers: a pilot study. *The Journal of Marketing, 47,* 68-78.

Richter, T. Schroeder, S. & Wohrmann, B. (2009) You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology, 96*, 538-558.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review, 5,* 296-320.

Sen, S. & Lerman, D. (2007) Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing, 21,* 76–94.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing, 21,* 76-94.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31,* 713-724.

Smith, D., Menon. S, & Sivakumar, K. (2005). Online Peer and Editorial Recommendations, Trust, and Choice in Virtual Markets. *Journal of Interactive Marketing, 19,* 15-37.

Trusov, M., Bucklin, R. E., & Pauwels, K. H. (2009). Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of Marketing, 73,* 90-102.

Vincenzi, M. E. (2012, September 2). Recensioni false e ricatti la crociata di ristoranti e hotel Denunciamo Tripadvisor. Retrieved from http://ricerca.repubblica.it/repubblica/archivio/repubblica/2012/09/02/recensioni-false-ricatti-la-crociata-di-ristoranti.html

Walster, E., & Festinger, L. (1962). The effectiveness of "overheard" persuasive communications. *Journal of Abnormal and Social Psychology, 65,* 395-402.

Webster, D. M., Richter, L., & Kruglanski, A. W. (1996). On leaping to conclusions when feeling tired: Mental fatigue effects on impressional primacy. *Journal of Experimental Social Psychology, 32,* 181-195.

Footnote

[1] Since having preliminary knowledge on a topic helps one to quickly recognise information as either true or false in that domain (Richter, Schroeder, & Wormann, 2009) we ran the same two-way analysis of variance including as a covariate the general judgment of online reviews' usefulness assessed at the beginning of our questionnaire. Indeed, we can suppose that individuals who considered the reviews useful read them frequently and are thus familiar with this online service. The analysis showed the main effect of the covariate factor, $F(1,74) = 12.51$; $p = .001$; $\eta^2_p = .14$, which however did not alter the described results.

Table 1.

Global evaluation of the restaurant (means, standard deviations in parentheses) as a function of falsity suspicion inducement and reviews valence order (N = 80).

Reviews valence order	Control condition	Suspicion condition	Total
Positive reviews first	3.58 (.76)	2.95 (.94)	3.27 (.91)
Negative reviews first	3.90 (.60)	2.17 (.92)	3.03 (1.17)
Total	3.74 (.70)	2.56 (1.00)	3.15 (1.04)

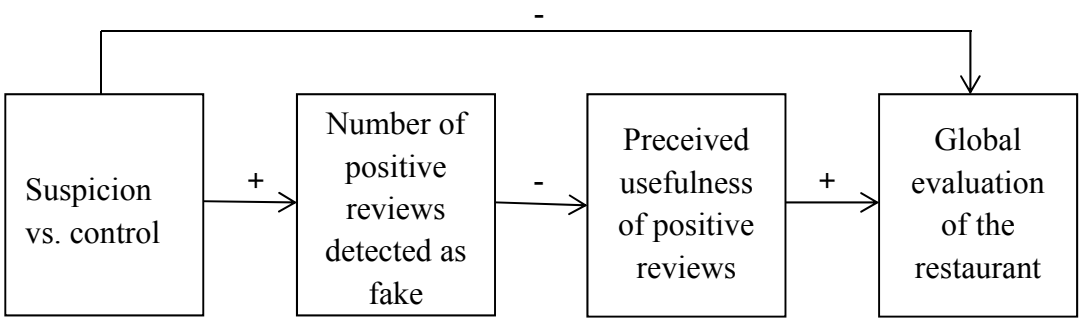

Figure 1.
Final mediational model of the falsity suspicion effect on restaurant evaluation.