Re-building Trust in Online Shops on Consumer Review Sites:

Sellers' Responses to User-Generated Complaints

Uwe Matzat   Chris Snijders

Sociology Section

Human-Technology Interaction Group

School of Innovation Sciences

Eindhoven University of Technology

The Netherlands

Version 31-07-2012

Re-building Trust in Online Shops on Consumer Review Sites:

Sellers' Responses to User-Generated Complaints

Abstract

How do online shops re-build trust on consumer-generated review sites after customers accuse them of misbehaving? Theories suggest that the effectiveness of responses depends on the type of accusation, yet online research indicates that apologies are superior to denials regardless of the type of accusation. We argue that customers are suspicious about online sellers, making denials implausible and ineffective in re-building trust. A good reputation may mitigate suspicion, making denials more believable and restoring trust. An experiment employed mock-ups of consumer review sites featuring different forms of consumers' complaints and shops' responses. Although reputable online shops were regarded as more trustworthy, results confirmed that denials tended not to be believed and did not re-build trust. Apologies generated superior effects.

Key words: trust, online shops, consumer review sites, trust restoration, reputation

Consumers' online feedback about products and services can be very influential. Numerous studies have shown that online consumer reviews influence consumers' risk perceptions and product choices (e.g., Senecal & Nantel, 2004). Obviously, not all consumer feedback is equally valued. In addition to other qualities (see for review Willemsen, Neijens, Bronner, & De Ridder, 2011), the valence of a user-generated review affects its usefulness and influence. Negative reviews have a strong impact on usefulness (Sparks & Browning, 2011; Willemsen et al., 2011), and they diminish other users' perceptions of a seller's trustworthiness (Pavlou & Dimoka, 2006). It is thus important for sellers to know how to react to a negative review. Anecdotal evidence suggests that providing no reaction to a consumer's negative review can be disastrous (Pantelidis, 2010). This article explores how a web shop that has been accused of misbehavior in a negative user-generated review can react successfully and restore trust.

Theories of trust restoration in face-to-face communication argue that the effectiveness of a reaction depends on the type of accusation. An *apology* restores trust more than a *denial* when a seller is accused of incompetence, whereas a denial is better when sellers are accused of immoral or opportunistic behavior (Kim, Ferrin, Cooper, & Dirks, 2004). In an online setting, matters might be different. Experimental research focusing on eBay's online feedback system has indicated that an apology is always trust-restoring, but not a denial, independent of the type of accusation (Utz, Matzat, & Snijders, 2009). It is an open question which conclusion pertains in other online settings such as online consumer review sites. Furthermore, it is unclear how to integrate the discrepancy between face-to-face versus online findings. Utz et al. (2009) offered one potential explanation. They argue that eBay users are distrusting towards eBay sellers and treat accused sellers as if their guilt has been established. This account can reconcile the divergent findings because the differential effects of apology vs. denial in face-to-face

communication emerge only under when guilt has not yet been established: A denial is counterproductive if guilt has been proven or is perceived to have been proven (Kim et al., 2004). In this article we test whether trust-restoration in online consumer review sites is more successful by means of an apology or a denial. Moreover, we test the potential explanation for the divergent prior findings by analyzing whether the effects suggested by Kim et al. (2004) also pertain to web shops that are regarded by consumers as highly reputable, for whom a denial may yet be a useful trust restoration response when a shop is accused of immoral behavior.

The Role of Trust in Electronic Commerce

Different forms of trust play an important role in stimulating buying behavior. Consumers' trust in systems such as the Internet (Grabner-Kräuter & Kaluscha, 2003), and institution-based trust in third-party mechanisms such as escrow services (Pavlou & Gefen, 2004) make e-commerce run smoothly. In addition, interpersonal trust—willingness to be vulnerable to a company that offers a specific product or service—is crucial in a customer-company relationship (Mayer, Davis, & Schoorman, 1995): One party hands control to the other party when one gives money to a web shop and must wait and see if the web shop will fulfill the contract as promised. A consumer's trusting disposition and the trustworthiness of the company both affect trust. The latter depends on the perceived ability, benevolence, and integrity, perceptions that pertain to the competence and the morality of a company (Mayer et al., 1995; Wojciszke, 2005). Belief in the trustworthiness of the other party makes acts of trust more likely.

Similarly, the trustworthiness of a web shop facilitates buying its products or services (Bhattacherjee, 2002). In addition, trust in web shops stimulates information disclosure by customers, who tend to have more trust in highly reputable web shops (Metzger, 2004; 2006), and for good reason: Research on online reputation systems indicates the aggregated score from

consumers' past feedback predicts future trustworthy behavior by sellers (Resnick & Zeckhauser, 2002). Moreover, a good reputation score increases trust (Snijders & Weesie, 2009) and leads to a small price-premium (Ba & Pavlou, 2002; Snijders & Zijdemann, 2004; see Cook, Snijders, Buskens, & Cheshire, 2009, for a general review of trust in online environments).

Trust Restoration in Face-to-Face versus Online Communication

Trust can be damaged in many ways, for instance, when one discovers intentional deception by the other party (Gregg & Scott, 2006). The restoration of trust, however, is not only important after an intentional violation of trust, but also in noisy environments where a breach of trust can easily appear to have happened even if it actually did not (Tazelaar, van Lange, & Ouwekerk, 2004). Noise is defined as "differences between actual and intended outcomes due to unintended errors" (van Lange, Ouwekerk, & Tazelaar, 2002, p. 768). Online communication between buyer and seller allows for several types of noise. One of the reasons is that the communication is physically separated and another is that there are simply several moments between initial contact and final delivery where things can go wrong. During transportation products may be damaged, employees may make mistakes resulting in a shipping delay, or shipments may go wrong completely. These mishaps can easily lead to negative reviews which can exert powerful effects on other potential customers (Sparks & Browning, 2011). Moreover, consumers tend to generalize their own negative experiences to the whole market (Pavlou & Gefen, 2005). It is therefore desirable for companies to re-build trust after they have received negative reviews.

It is unclear how a company can respond best to negative online reviews. Theory and research on trust re-building in (early phases of) face-to-face relationships suggest one set of contengencies. Findings about trust-restoration between buyers and sellers on eBay, however,

show remarkable differences in the effectiveness of trust restoration strategies. A potential explanation for these differences that could contribute to an integration of the diverging findings leads to a number of testable predictions that appear at the end of this section.

Face-to-face communication. Studies on trust restoration in early phases of face-to-face relationships examined the value of two broad strategies, namely *apologizing* (acknowledging both responsibility for and regretting a trust-violation) versus *denying* (declaring an accusation to be untrue). Various studies differ, however, with respect to the effectiveness of each approach (see for review Schweitzer, Hershey, & Bradlow, 2006). Kim et al. (2004) reconciled the divergent findings on the strategies' utility by assuming that their effectiveness depends on which of two types of trust violation that a party has been accused. Individuals can accuse another party of an integrity-based, or *morality-based violation of trust*, suggesting that the other party has knowingly behaved in a way that does not adhere to generally accepted principles. Individuals can also accuse another party of not having the required skills to deliver the contracted good or service properly, which we label as a *competence-based violation of trust*.

Kim et al. (2004) argue that, as long as an accused's guilt has not been established, a denial would be more successful in re-building trust following a morality-based accusation. In contrast, an apology would be more successful in case of a competence-based trust violation. Kim et al. base this claim on insights from research on causal inference and interpersonal perception (Reeder & Brewer, 1979). This research demonstrates that in the eyes of observers negative information about other parties in the *domain of morality* (or integrity) issues is considered more informative about the parties' presumable traits than is positive information. At the same time, positive information about another individual is more diagnostic than negative information when it comes to inferring a party's traits in the *domain of skills and abilities*

(Skowronski & Carlston, 1987). This asymmetry occurs because humans use so-called "hierarchically restrictive schema" for inferring dispositions from somebody's behavior (Reeder & Brewer, 1979): Humans believe that moral as well as immoral individuals can show moral behavior from time to time, whereas only an immoral person would show immoral behavior. In this sense, immoral behavior implies revealing oneself as being the immoral type (Skowronski & Carlston, 1987). The situation is reversed when observers make causal inferences about dispositions in the domain of skills and abilities. Individuals with little skill would be more restricted in their behavior than individuals with greater skill (Reeder & Brewer, 1979). A skilled individual may deliver a weak performance because of a lack of motivation or bad luck. Good performance, however, must necessarily be attributed to a strong skill level because individuals with few skills are simply unable to perform well (Skowronski & Carlston, 1987).

As long as guilt has not been proven, we argue, when a third party sees a morality-based accusation of a trust violation (accusing the seller as having knowingly offended, e.g., "You deliberately did not send me my package in time"), an apology by the alleged trust breaker acknowledges, at least implicitly, that the accusation is correct. This reveals negative information about the accused party's morality disposition, leading to a decline in observers' perceptions of the morality of the accused party, and as a result, its trustworthiness. A denial avoids revealing negative information about the accused party's morality—the alleged trust-breaker might be innocent—making a denial a more effective trust-rebuilding strategy than an apology in the case of a morality-based accusation (Kim et al., 2004). In case of a competence-based accusation of a trust violation, an apology reveals negative information, which is not diagnostic in the competence domain, causing no serious reduction in the perceived competence of the other party and therefore leaving the perception of trustworthiness unchanged. In addition, an apology

signals intended redemption. In contrast, a denial would not lead to a reduction of perceived trustworthiness, but at the same time would eliminate any signal of intended redemption, making an apology the more effective trust re-building strategy in the context of a competence-based user-generated accusation (Kim et al., 2004).

Kim et al. (2004) tested the predictions in a laboratory experiment with undergraduate students who watched a videotaped accounting firm job interview. During the interview an applicant was accused of having incorrectly filled in a tax return form at her last job. After watching the video the students rated the applicant's trustworthiness. The experimental procedure randomly assigned participants to one of the four conditions that crossed the two types of violations and two types of reactions. As long as the interviewee's guilt had not been proven, the hypotheses were supported, but not when guilt was established, as the authors had expected.

Online communication. Through poor numerical feedback in online reputation systems (see, e.g., Snijders & Zijdeman, 2004) or through narrative online reviews, user-generated content can accuse a company of violations of trust and can thereby reduce the company's trustworthiness among others (Pavlou & Dimoka, 2006; Pavlou & Gefen, 2005, Sparks & Browning, 2011). There is some evidence indicating that it is possible for sellers to rebuild trust, at least on eBay: Utz (2009), Utz et al. (2009), and Bober et al. (2011) investigated trust re-building effects of apologies versus denials under the two types of accusations of trust violations. They randomly assigned eBay users to view hypothetical profiles of sellers. Apart from numerical information about the seller's past evaluation that was kept constant, the profiles included different versions of textual feedback that corresponded to the four combinations of buyers' accusations and sellers' reactions. Every profile included one negative user-generated complaint with a competence- or morality-based accusation by a former buyer of that seller. In

the stimulus profile the seller either did not react, offered an apology, or denied the accusation. The participants' subsequent ratings of the seller's trustworthiness showed that morality-based accusations reduced trustworthiness more than competence-based accusations. An apology rebuilt trustworthiness completely under both types of accusations, whereas a denial under both types of accusations did not (Bober et al., 2011, Utz et al., 2009). Furthermore, trust restoration was mediated by the believability of the reaction, a strong predictor of perceived trustworthiness (Utz et al., 2009). The stronger effect of apologizing under both conditions departs from the findings of Kim et al. (2004) about trust re-building in face-to-face relationships.

Hypothetical Explanation of the Divergent Findings

Does this mean that the mechanisms of trust-rebuilding are different for online vs. face-to-face communication? Utz et al. (2009) offered the following explanation for their unexpected finding. First, they rejected the possibility that participants did not comprehend the written nature of the comments in the electronic medium. Utz et al.'s (2009) results demonstrated that both accusations reduced trustworthiness (see Bober et al., 2011) and were also correctly understood, just as the sellers' reactions were. Denials were less believable than apologies were: If eBay customers do not believe accused sellers' denials, they treat them as if guilt has been established. Under the condition of established guilt, Kim et al.'s (2004) theory no longer predicts that a denial has a stronger effect than an apology as a reaction to a morality-based accusation. Indeed, other research has shown that eBay users, when compared to users of several other online communities, show rather low levels of trust (Matzat, 2009).

This potential explanation, that eBay users are likely to be suspicious about other eBay sellers, making a denial unbelievable and therefore less effective as a trust-rebuilding strategy, could reconcile the contrasting findings of Kim et al. (2004) and Utz et al. (2009). The

explanation, however, was formulated post-hoc and could not be tested in previous studies. The following study provides an original test of this explanation. Our approach is based on the idea that, in principle, one would need to compare web sites that people tend to trust more with those that people tend to trust less. In absence of eBay-like websites that are highly trusted, we turned to a comparison of online shops.

We expect customers of web shops to believe the accusations that are generated and posted online by other users. Given that negative information is more diagnostic in the domain of morality we expect that a user-generated morality-based accusation reduces seller's trustworthiness more than a competence-based accusation. In addition, an individual's trusting disposition is known to influence trustworthiness (Mayer, Davis, & Schoorman, 1995). This leads to the following hypotheses.

H1: The trustworthiness of a web shop that has been accused of a morality-based violation of trust by user-generated content is less than the trustworthiness of an online shop that has been accused of a competence-based violation of trust.

H2: The greater the trusting disposition of an individual is, the more trustworthiness he or she perceives for an online shop.

In general customers do not believe a denial posted by a web shop following a user-generated accusation of a morality-based trust violation. Because readers treat a web shop as if guilt has been established they believe an apology more than a denial. Accordingly, a denial posting does not restore trust after a morality-based accusation, but an apology does. In addition, an apology is also a superior trust restoration strategy than a denial after a competence-based user-generated accusation. This leads to the following two hypotheses.

H3: When a web shop receives a user-generated accusation of a trust violation, customers

believe an apology by the web shop more than a denial (independent of the kind of accusation).

H4: An apology that follows a user-generated accusation of a trust violation increases customers' trust in the web shop (compared to not reacting to the accusation).

Although customers may not believe denials by accused web shops in general, there may be exceptions. Many web shops are small and relatively unknown to customers, but others are established and enjoy a good reputation. They are large companies, and/or they use official third-party certificates to signal trustworthiness (Jarvenpaa, Tractinsky, & Saarinen, 1999; Jiang, Jones, & Javie, 2008; Kim & Kim, 2011; Metzger, 2006). For well-trusted web shops, it may not be the case that a user-generated accusation leads other customers to perceive the shop as guilty and to disbelieve the shop's reactions. If so, these web shops would do better to react to a morality-based trust violation by denying it.

H5: As a reaction to a morality-based user-generated accusation, customers tend to believe a denial from a highly reputable web shop more than a denial from a less reputable web shop.

H6: A denial reaction to a morality-based accusation of trust violation increases customers' trust more for a highly reputable web shop than it does for a less reputable web shop (as compared to not reacting to an accusation at all).

Method

A sample of 322 Dutch users of online review sites were recruited in October 2011 via a commercial opt-in research participation panel. Among participants who completed the survey, we excluded the data from 13 participants who completed the experiment in less than five minutes (within which it is hardly possible to read the online material). We also disregarded the

cases of 20 other participants whose data appeared suspicious for other reasons, such as an excessive number of the same answers in matrix questions. This left 289 participants who completed the questions in a median time of 13 minutes (with 90% of the participants using between 7 and 34 minutes). All analyses are based on these $N = 289$ participants. The average age of participants was 45 years ($SD = 14.5$, range 18-78 years). Half of the participants (49 %) were male, and 9 % had an academic education. About half of the participants (47 %) had been using the internet for 10 years or longer. The participants bought an average of almost three products on the internet during the last 12 months (range from one to more than 20 products). More than 60 % of the participants judged online shops in general as reasonably reliable or better, 6.5 % found them hardly reliable or worse, and the other 34% were more moderate. All participants had used an online review site before. These numbers suggest a heterogeneous population of internet users with some experience in online shopping and familiarity with online review sites.

When the experiment started, instructions asked participants to imagine that they were looking for a specific digital camera that would cost about 200-250 Euros in different online shops. The experiment system presented a screenshot of a hypothetical consumer review site that showed a negative evaluation from a former buyer of the camera under consideration, followed (in most cases) by an online response from the shop. The screenshots of the consumer review site varied along several dimensions, including an online shop of weak vs. strong reputation, a user-generated accusation that was either morality- or competence-based, and the kind of reactions posted by the online shop: no reaction, plain apology, extended apology, plain denial, or extended denial, in accordance with Utz et al. (2009). Moreover, to enhance generalizability of the findings, we used two different hypothetical incidents of different severity. One incident

described a delayed delivery of the camera whereas the other incident considered the delivery of a camera with some defect, two of the more common matters that can go wrong in an online purchase. Taken together, this implies we had a 2 x 2 x 5 x 2 design which yielded 40 different scenarios. After studying one of these presentations, participants answered a number of questions about their perceived trustworthiness of the online shop, the believability of the customer's accusation, and the believability of the web shop's reaction. Each participant viewed and responded to a random set of eight of these scenarios; four featured the broken product incident, and four featured the delayed delivery incident. This makes the arrangement a "mixed design" (cf. Field, 2009) with eight observations per participant. Finally, the participants answered questions about their trusting disposition, attitudes towards online shopping, and demographics.

Independent Variables

Type of incident. Two different incidents were presented to the participants. The first was a broken-product scenario, in which the user-generated complaint indicated that the camera which the user had purchased from the web shop was broken upon delivery. The second version presented a delayed-delivery scenario: The camera arrived 6 weeks later than had been promised.

Reputation of online shop. The web shops' reputations were varied in order to instill greater vs. lesser trust. The upper part of the review page displayed the name of the online shop together with some background information about its history, web site address, and telephone number. It also displayed an aggregated numerical score that represented the evaluations of other users. The strong reputation condition depicted the online shop Bol.com, which is, in reality, one of the largest and best known Dutch online shops; it has strong market penetration and has won several annual "best service to clients" awards (http://tiny.cc/l5vbhw). On the review site its score indicated that 93% of former customers had recommended the shop, that the

online shop existed since 1999, and that it recently won a web shop award. Moreover, two third-party certificates were presented in a prominent place. Alternatively, in the weak reputation condition, an unknown (non-existing) web shop called ElektroDiscount.com appeared. Its recommendation score was 63%, its background information included only some brief contact information, and it featured no third-party certificates. A manipulation check verified the expected difference in perceived trustworthiness between the two types of shops. The strong reputation condition's (Bol.com) trustworthiness was $M = 5.3$ ($SE = 0.04$), whereas the trustworthiness of ElektroDiscount was $M = 4.1$ ($SE = 0.04$), $t(2310) = 21.5$, $p < .001$.

Type of trust violation. The user-generated complaint indicated that the web shop had committed either a morality-based trust violation or a competence-based violation. In the broken-product scenario, the morality-based violation of trust was reflected by the following customer accusation (translated from Dutch): "*The camera I received is broken. The wrapping had already been removed; that's cheating!*" whereas the competence-based violation of trust was represented by "*The camera I received is broken. It seems that the shop was sloppy while packaging; there was far too little foam synthetics around it.*" In the delayed delivery scenario, the accused morality-based trust violation was connoted with the comment, "*Unreliable shop! The camera was sent 6 weeks later than promised!*" whereas the competence-based violation was indicated by "*Sloppy shop: they sent it to the wrong postal code! Took 6 weeks before I received the product.*" These manipulations have been used in previous studies in which it was shown that users understand the difference between the two types of trust violation (for review, see Bober et al., 2011).

Type of posted reaction. The shops appeared to have posted either an apology for or a denial of the accused trust violation, and the content of these reactions depended on the type of

scenario and type of accusation. Moreover, to increase the generalizability and to depend less on the particular wording of single comments, every reaction by an online shop was replicated in a plain version that was very short and did not offer a potential explanation of what had happened, and an extended version that did offer a potential explanation. For instance, in the broken-product scenario with a morality-based trust violation, the plain denial was "*We understand that the situation is unpleasant for you. However, this was not our mistake. We never open products,*" whereas the extended denial was "*We understand that the situation is unpleasant for you. However, this was not our mistake. We never open products. Maybe something went wrong during the shipment?*" Pre-tests and findings in other studies have shown that users understand the difference between the two types of reactions (Bober et al., 2011). A complete overview of all reaction postings appears in Appendix A.

Dispositional Trust

Dispositional trust was measured with the 4-item, five-interval Likert scales developed by Jarvenpaa, Knoll, and Leidner (1998). Examples include, "Most people are honest in describing their experiences and abilities," and "Most people are honest in describing their experiences and abilities." Cronbach's α = .84.

Dependent Variables

Perceived trustworthiness of the seller was measured with seven-interval Likert scale items based on the measure by McKnight, Choudhury, and Kacmar (2002), adapted by Utz et al. (2009) to the context of online shops. Items included, "This online shop is competent," and "I think this shop wants the best for the buyers," α = .98. The perceived believability of the shop reaction was measured with the single seven-interval Likert item, "I consider the reaction of the shop as highly believable,"

Method of Data Analysis

The presentation of eight scenarios per participant creates dependence that should be accounted for in the analyses (Bryk & Raudenbush, 1992). This can be done using multilevel regression models (also called hierarchical models) for multivariate data analyses (see, e.g., Goldstein, 1995; Leckie, 2010), a variant of repeated measures analyses or mixed models. Because the eight cases per participant do not deliver independent observations, standard confidence intervals from analysis of (co-) variance or multiple regression would likely be too narrow. Multilevel analysis accounts for these (and some other) issues. We refer the reader to Snijders and Bosker (2011) for further details.

<div align="center">

Results

</div>

Several potential covariates did not approach significance and are not included in further analyses: participant sex, previous purchase from Bol.com in the past 12 months, level of education, time online per week, assessment of online shops' general reliability, and whether photography was a participant's hobby (assessed because the product in the experimental scenarios was a digital camera).

The mean level of perceived trustworthiness in the target web shop was 4.7 (*SE* = 1.4) out of 7. As mentioned above, there was a sizeable difference between the trust in Bol.com vs. ElektroDiscount in accordance with the intended manipulation. In addition to the models we present below, we examined interaction effects and different parameters for retaining or removing participants whose responses raised suspicions of invalid data entry; none of these altered the major results.

In order to assess the necessity of using multilevel analysis, initial diagnostics showed that the variance at the individual level is substantial: An empty variance component model

indicates that 24% of the variance in the trustworthiness ratings resides with the individual. Adding individual characteristics to this model that we included in our data collection as possible covariates (age, trust disposition, how reliable the participant considered Bol.com, and the ease with which participants spend money on gadgets and electronics) decreases this to 21%. No statistically significant interactions of any of these covariates with experimental factors existed across all our models (all $p > 0.3$), which confirms the viability of multilevel modeling.

We now consider several models that we estimated to test our hypotheses (see Table 1), with perceived trustworthiness of the seller as the dependent variable. All models include the individual characteristics mentioned above. In model I, we show a baseline model with 8 observations per $N = 289$ subjects, or 2,312 cases that includes whether there was a morality accusation, which scenario was used (broken vs. late camera), which shop was being evaluated, and whether an apology or denial was offered. In Model I, we use a single variable "apologies" for both the simple and the extended apology, and a single variable "denials" for both the simple and the extended denial. Models II and III show the results for the different scenarios separately which leads to a halving of the N of cases per model ($N = 1,156$). Model II presents the results for the late delivery scenario, whereas Model III presents the results for the broken-product scenario. In Models IV and V, we rerun Models II and III, adding explicit distinctions between simple and extended denials, and simple and extended apologies. Finally, Model VI takes into account only morality accusations (as indicated in H6) and includes the interaction effect of the type of shop by apologies and denials. In this final model, both scenarios are collapsed which again leads to an N of cases = 1,156. Taken together, these model estimates allow a test of our hypotheses, as outlined below.

The results show no evidence that morality accusations lead to less perceived

trustworthiness, refuting H1. This can be seen from the coefficient of the variable "morality accusation," which is not significant in any of the models (and restricting the sample to only those cases without a reaction from the shops does not change that result). From the significant coefficient for the variable "trust disposition" across all models we can conclude that as the general trusting disposition of a participant is greater, the participant also trusts both the shops more, supporting H2. In addition, throughout the analyses, there is a positive main effect of the perceived reliability of Bol.com on trustworthiness of the web shop. In addition, irrespective of whether we consider the apologies (simple versus extended) independently or together, we find that apologies increase the participants' trust in the online shops compared to seeing no reaction postings by the shops at all; the coefficients for "comment is apology," "apology plain," and "apology extended" are all significant and positive. This supports H4.

Interaction effects between any type of reaction posted by the shops and the type of accusation did not reach significance (the smallest $p = .19$). Although we did not have particular hypotheses about the (absolute) effect of denials, we found no positive effect of denials on trust. In fact, in the broken product scenario, the plain denial actually decreased trust in the shop, whereas in all other cases the denials had no effect on trust. Finally, model VI's results refute H6: It is not the case that a denial posting from the more reputable shop increases trust more than a denial from a less reputable shop (i.e., there is a no significant effect of the variable "reputable store X denials" in model VI). Additional analyses that only include the participants with more trusting dispositions do not change this result. We will return to the interpretation of this result later. An interesting and unexpected result is that apologies increase trust, but *less* so for the more reputable shop (a negative and significant effect of "reputable store X apologies" in model VI).

(Table 1 here)

We now turn to the analyses of the believability of the reaction generated by the web shop as the outcome variable. Before we describe the hypotheses tests on believability in Table 2, it is important to note that there is an overall difference in the direction of the believability of reactions from the less reputable shop (M = 4.2, SE = 0.05) and the more reputable shop (M = 4.9, SE = 0.05), t (1857) = 10.7, p < .001. This difference remains significant and of the same magnitude when we consider only the denials (M = 3.7, SE = 0.07 vs. M = 4.4, SE = 0.07), t (896) = 7.2, p <.001). Believability of apologetic postings also differed due to whether the shop had a weaker reputation (M = 4.7, SE = 0.06) or a stronger one (M = 5.3, SE = 0.05), t (959) = 8.3, p <.001) despite the main effect of apologies vs. denials. This suggests that, although the more reputable firm is more believable irrespective of the kind of response it offers, denials do not seem to overcome these differences. Once again, several models test the hypotheses while controlling for other covariates, the results of which appear in Table 2. In model VII we use the same base model as in model I: a single variable, "apologies," for both the simple and the extended apology, and a single variable, "denials," for both the simple and the extended denial, and no differentiation between scenarios. Because believability of the reaction is the outcome variable, the cases where no reaction was presented were excluded from the analyses this time, leading to 1,859 cases for Model VII. From the negative coefficient of the variable "Reaction is denial" (vs. apology or nothing) we conclude that apologies are indeed more believable than denials, in accordance with H3. Two subsequent models add the interaction effect of whether the store is a reputable one with whether the reaction is a denial. Model VIII uses the cases where the morality-based accusation appeared, and model IX considers the competence-based accusations. Hypothesis 5 suggested that the effect of a denial on believability should be greater

for the more reputable web shop, but the results do not support this assertion for the morality-based accusation with which H5 was concerned, nor for the competence-based accusation (i.e., the coefficient of "reputable store X reaction is denial" is not significant in models VIII and IX).

(Table 2 here)

Finally, when we combine the analyses from Table 1 and 2, and include believability of the reaction as a covariate for the outcome variable of trust in the shop, in Models I through VI, believability is a strong predictor in all models ($p < .001$; all z-values > 30). Hence, the results indicate that believability of the shops' reaction postings plays a strong role when it comes to customers' assessment of trust in those shops. Nevertheless the reaction messages from the reputable shop are more believable, and the reputable shop, in general, is trusted more than the non-reputable shop. What is *not* supported in the data is the idea that shops differ in the extent to which a denial can restore trust (as compared to not reacting at all). A less reputable shop does not restore trust by denying an accusation as predicted, but, contrary to our expectation, the same goes for a more reputable shop, even though (after a denial) the residual trust in the more reputable shop is still greater than in the less reputable shop.

To summarize the findings, the data support some of the hypotheses, but not all. Apart from the trust-increasing effect of the user's trusting disposition (H2), users believe apologies more than denials (H3). Moreover, an apology re-builds trust (H4), whereas a denial does not. However, there are no differences due to the the two types of accusations (H1), nor is there a difference in the believability (H5) and trust restoring effect (H6) of a denial due to the prior reputation of a web shop. These findings have clear implications for trust restoration strategies of web shops on Web 2.0 consumer review sites. Furthermore, there are theoretical implications for the integration of the (seemingly) divergent online vs. offline findings on the mechanisms of

trust re-building, as we elaborate below.

Discussion

Because negative user-generated reviews in online review sites can seriously harm consumers' trust in web shops, questions arise whether and how online shops can restore trust after they have been accused of having misbehaved. Previous research revealed a disparity between the sellers' response strategies that are most successful in traditional, face-to-face settings, compared to those in a study of eBay. Although a seller's denial restores trust following an accusation of intentional wrongdoing offline (Kim et al., 2004), denials were less effective than apologies online (Utz et al., 2009). Potential buyers may not believe denials by an eBay seller, and assume that eBay sellers' guilt is a foregone conclusion. This interpretation resolves the apparent contradiction between face-to-face and online research. The present research tested this potential theoretical reconciliation in a field experiment among Dutch users of online review sites. It also examined the hypothesis that highly reputable shops are immune to suspicion about online shops in general. If so, customers may believe a denial by a reputable web shop more than for less well-regarded web shops, and a reputable seller's denial should restore trust when faced with a morality-based accusation. Findings indicate that when a site user posts an accusation that a web shop seriously erred—intentionally or accidentally—the seller may rebuild trust by certain kinds of online replies. An apology restores trust successfully, whereas a denial does not and may even reduce trust, consistent with the earlier findings of Utz et al. (2009). This seems to be because customers believe a denial much less than an apology. Contrary to our expectations, even though customers found the denials by a reputable web shop more believable than denials by a less reputable web shop, denials of morality-based accusation (i.e., an intentional error) by a highly reputable web shop restores trust no more than a denial by less reputable web shops. An

apology by a web shop is always more successful in re-building trust than a denial, under both types of accusations and regardless of reputation.

One contribution of this research is that it clarifies the underlying mechanism—the presumption of guilt—as a factor that leads to users' acceptance or rejection of a seller's denial of intentional misbehavior. The poor believability that participants attributed to web shops' denials makes this explanation more likely to be true. It may be that customers tend to believe an apology that accepts responsibility for having misbehaved, but not a denial that does not accept responsibility. This presumption is likely to invoke the same mechanisms that affect trust restoration in face-to-face communication: the assumption of established guilt, which undermines a denial's effectiveness in traditional communication settings (Kim et al., 2004). However, our results offer another possible explanation that can integrate the divergent findings. It could be that customers equate *believability* with *showing responsibility*. For the average customer it seems to be part of the web shop's responsibilities to choose competent and honest personnel and a competent and honest delivery service. Buyers may consider that all responsibility for a successful transaction are the online seller's burden.

Should this be true, no denial may be plausible since all problems are the seller's responsibility. Only an acceptance of responsibility (by way of apology) can be persuasive. This interpretation strengthens the notion that presumption of guilt operates more consistently with regard to web shops, reinforcing the principle articulated by Kim et al. (2004). This explanation is also consistent with the finding that the expected difference on trust ratings due to a morality-vs. a competence-based accusation did not occur for the web shops. What may matter, then, is *responsibility*, not intention or guilt in the actual sense.

This study has limitations that should be addressed in future research. First, the

explanation we offered post-hoc for the refutation of H5 and H6 cannot be tested with our data; future experiments with an explicit focus on this issue should do so. It may be that other types of denials, that do not shift responsibility away from the web shop, re-build trust more successfully. A shop may explicitly accept responsibility without addressing whether the mistake was or was not theirs. Another limitation of our study is that it only analyzes the effects of single iterations of online accusations and reactions,. It would be worthwhile to study whether the effectiveness of the strategies changes when user-generated content contains more accusations, a combination of positive and negative evaluations, corroborations or dismissals in other users' comments, and whether other kinds of strategies (such as evasion or compensation) have effects on trust as well. Also, there is great variety among user-generated comments in terms of how strong or how negative the wording is, not to mention potential effects of the numerical scores describing the sellers as well as the commenters, which may all affect the strategies' effectiveness.

Because apologies are effective at re-building trust but denials are not,web shops face a dilemma when they are falsely accused. Even in such cases a denial does not restore trust. Web shops may have to realize that customers hold them accountable for incidents that are beyond their control. The dilemma of web shops is even more severe when one remembers another of Kim et al.'s (2004) findings: Those who apologized for something they did not do increased customers' trust only as long as guilt has not been established. Apologizing for something one has not done, however, is counterproductive if it turns out be an overt lie. It may therefore be wise for web shops to acknowledge responsibility for negative incidents caused by third parties but to do so without conceding that the offending actions are under their immediate control. By doing so, web shops ensure authenticity of their behavior and prevent being perceived as disingenuous at some later time.

References

Ba, S. L., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in

    electronic markets: Price premiums and buyer behavior. *MIS Quarterly, 26*, 243-268.

    doi:10.2307/4132332

Bhattacherjee, A. (2002). Individual trust in online firms: Scale development and initial test.

    Journal of Management Information Systems, 19, 211-241. doi:10.4018/jebr.2006100102

Bober, M., Snijders, C., & Matzat, U. (2011). *User authority in productive consumption on

    eBay: Damaging and rebuilding trust through feedback comments*. Manuscript submitted

    for publication.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data

    analysis methods*. Newbury Park, CA: Sage.

Cook, K., Snijders, C., Buskens, V. & Cheshire, C. (2009). *eTrust: Forming relationships in the

    online world*. New York, NY: Russell Sage Foundation.

Field, A. (2009) *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA: Sage.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.) London, England: Edward Arnold.

Grabner-Krauter, S., & Kaluscha, E. A. (2003). Empirical research in on-line trust: A review and

    critical assessment. *International Journal of Human-Computer Studies, 58,* 783-812.

    doi:10.1016/S1071-5819(03)00043-0

Gregg, D. G., & Scott, J. E. (2006). The role of reputation systems in reducing on-line auction

    fraud. *International Journal of Electronic Commerce*, *10*(3), 95-120.

    doi:10.2753/JEC1086-4415100304

Jarvenpaa, S. L., Knoll, K., & Leidner, D. E. (1998). Is anybody out there? Antecedents of trust

    in global virtual teams. *Journal of Management Information Systems*, *14*, 29-64.

    Retrieved from http://www.jstor.org/discover/10.2307/40398291?uid=3739728&uid=2&

uid=4&uid=3739256&sid=47699131101037

Jarvenpaa, S. L., Tractinsky, N., & Saarinen, L. (1999). Consumer trust in an internet store: A

cross-cultural validation. *Journal of Computer-Mediated Communication, 5*(2).

doi:10.1111/j.1083-6101.1999.tb00337.x

Jiang, P. J., Jones, D. B., & Javie, S. (2008). How third-party certification programs relate to

consumer trust in online transactions: An exploratory study. *Psychology & Marketing,

25,* 839-858. doi:10.1002/mar.20243

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of

suspicion: The effects of apology versus denial for repairing competence- versus

integrity-based trust violations. *Journal of Applied Psychology, 89,* 104-118.

doi:10.1037/0021-9010.89.1.104

Kim, K., & Kim, J. (2011). Third-party privacy certification as an online advertising strategy: An

investigation of the factors affecting the relationship between third-party certification and

initial trust. *Journal of Interactive Marketing, 25,* 145-158.

doi:10.1016/j.intmar.2010.09.003

Leckie, G. (2010). *Stata practical: Introduction to multilevel modeling.* Centre for Multilevel

Modeling. Retrieved from http://www.bristol.ac.uk/cmm/learning/module-samples/5-

practicals-stata-sample.pdf

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. doi:10.1016/j.elerap.2007.05.004,

Matzat, U. (2009). The acceptance and effectiveness of social control on the Internet: A

comparison between online auctions and knowledge sharing groups. In K. S. Cook, C.

Snijders, & V. Buskens (Eds.), *eTrust: Forming relationships in the online world* (pp.

238-265). New York, NY: Russell Sage Foundation.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational

trust. *Academy of Management Review, 20,* 709-734. doi:10.2307/258792

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust

measures for e-Commerce: an integrative typology. *Information Systems Research, 13*,

334-361. doi:10.1287/isre.13.3.334.81

Metzger, M. J. (2004). Privacy, trust, and disclosure: Exploring barriers to electronic commerce.

Journal of Computer-Mediated Communication, 9(4). doi:10.1111/j.1083-

6101.2004.tb00292.x

Metzger, M. J. (2006). Effects of site, vendor, and consumer characteristics on web site trust and

disclosure. *Communication Research, 33,* 155-179. doi:10.1177/0093650206287076

Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant

comments. *Cornell Hospitality Quarterly, 51,* 483-491. doi:10.1177/1938965510378574

Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based

trust. *Information Systems Research, 15,* 37-59. doi:10.1287/isre.1040.0015

Pavlou, P. A., & Gefen, D. (2005). Psychological contract violation in online marketplaces:

Antecedents, consequences, and moderating role. *Information Systems Research, 16,* 372-

399. doi:10.1287/isre.1050.0065

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online

marketplaces: Implications for trust building, price premiums, and seller differentiation.

Information Systems Research, 17, 392-414. doi:10.1287/isre.1060.0106

Reeder, G. D., & Brewer, M. B. (1979). Schematic model of dispositional attribution in

interpersonal perception. *Psychological Review, 86,* 61-79. doi:10.1037/0033-295X.86.1.61

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In M. R. Baye (Ed.), *The economics of the internet and e-commerce* (pp. 127-157). Bingley, England: Emerald Group Publishing Limited.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101,* 1-19. doi:10.1016/j.obhdp.2006.05.005

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing, 80,* 159-169. doi:10.1016/j.jretai.2004.04.001

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52,* 689-699. doi:10.1037/0022-3514.52.4.689

Snijders, C., & Weesie, J. (2009). Trust and reputation in an online programmer's market: A case study. In K. S. Cook, C. Snijders, V. Buskens, & C. Cheshire (Eds.), *Trust and reputation* (pp. 166-185). New York, NY: Russell Sage Foundation.

Snijders, C., & Zijdeman, R. (2004). Reputation and Internet auctions: EBay and beyond. *Analyse & Kritik, 26,* 158-184. Retrieved from http://igitur-archive.library.uu.nl/fss/2011-1208-200424/AK_Snijders_2004.pdf

Snijders, T. A. B., & Bosker, R. J. (2011) *Multi-level analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, England: Sage.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions

and perception of trust. *Tourism Management, 32,* 1310-1323.

doi:10.1016/j.tourman.2010.12.011

Tazelaar, M. J. A., Van Lange, P. A. M., & Ouwerkerk, J. W. (2004). How to cope with "noise"

in social dilemmas: The benefits of communication. *Journal of Personality and Social*

Psychology, 87, 845-859. doi:10.1037/0022-3514.87.6.845

Utz, S., (2009). Rebuilding trust after negative feedback: the role of communication. In K. S.

Cook, C. Snijders, & V. Buskens (Eds.), *eTrust: Forming relationships in the online*

world (pp. 238-265). New York, NY: Russell Sage Foundation.

Utz, S., Matzat, U., & Snijders, C. (2009). Online reputation systems: The effects of feedback

comments and reactions on building and rebuilding trust in online auctions. *International*

Journal of Electronic Commerce, 13, 95-118. doi:10.2753/JEC1086-4415130304

Van Lange, P. A. M., Ouwerkerk, J. W., & Tazelaar, M. J. A. (2002). How to overcome the

detrimental effects of noise in social interaction: The benefits of generosity. *Journal of*

Personality and Social Psychology, 82, 768-780. doi:10.1037/0022-3514.82.5.768

Willemsen, L. M., Neijens, P. C., Bronner, F., & De Ridder, J. A. (2011). Highly recommended:

The content characteristics and perceived usefulness of online consumer reviews. *Journal*

of Computer-Mediated Communication, 17, 19-38. doi:10.1111/j.1083-

6101.2011.01551.x

Wojciszke, B. (2005). Affective concomitants of information on morality and competence.

European Psychologist, 10, 60-70. doi:10.1027/1016-9040.10.1.60

Appendix A

Broken product scenario (All texts were originally in Dutch.)

Condition	Customer accusation	Shop reaction
Accusation of incompetence + no reaction		
Accusation of incompetence + plain denial		We understand that the situation is unpleasant for you. However, this was not our mistake. We have used this product packaging since years and never had any problems.
Accusation of incompetence + extended denial	The camera I received is broken. It seems that the shop was sloppy while packaging; there was far too little foam synthetics around it.	We understand that the situation is unpleasant for you. However, this was not our mistake. We have used this product packaging since years and never had any problems. Maybe something went wrong during the shipment?
Accusation of incompetence + plain apology		Our apologies. We will forward your complaint to the logistics unit and will look for a solution.
Accusation of incompetence + extended apology		Our apologies, we use new packaging material. We will forward your complaint to the logistics unit and will look for a solution.
Accusation of immorality + no reaction		
Accusation of immorality + plain denial		We understand that the situation is unpleasant for you. However, this was not our mistake. We never open products.
Accusation of immorality + extended denial	The camera I received is broken. The wrapping had already been removed; that's cheating!	We understand that the situation is unpleasant for you. However, this was not our mistake. We never open products. Maybe something went wrong during the shipment?
Accusation of immorality + plain apology		Our apologies, we are sorry for this, this never should have happened. We are looking for a solution.
Accusation of immorality + extended apology		Our apologies, we are sorry for this, this never should have happened. Probably our controllers missed this. We are looking for a solution.

Delayed delivery scenario (All texts were originally in Dutch.)

Condition	Customer accusation	Shop reaction
Accusation of incompetence + no reaction		
Accusation of incompetence + plain denial		We understand that the situation is unpleasant for you. However, this was not our mistake. We sent the product with the correct postal code.
Accusation of incompetence + extended denial	Sloppy shop: they sent it to the wrong postal code! Took 6 weeks before I received the product.	We understand that the situation is unpleasant for you. However, this was not our mistake. We sent the product with the correct postal code. Maybe something went wrong at the carrier?
Accusation of incompetence + plain apology		Our apologies, we are sorry for this, this never should have happened. We will intensify our control procedure so that this does not happen again.
Accusation of incompetence + extended apology		Our apologies, we are sorry for this, this never should have happened. This happened because of a new employee. We will intensify our control procedure so that this does not happen again.
Accusation of immorality + no reaction		
Accusation of immorality + plain denial		We understand that the situation is unpleasant for you. However, this was not our mistake. We gave the product in time to the carrier.
Accusation of immorality + extended denial	Unreliable shop! The camera was sent 6 weeks later than promised!	We understand that the situation is unpleasant for you. However, this was not our mistake. We gave the product in time to the carrier. Maybe something went wrong at the carrier?
Accusation of immorality + plain apology		Our apologies, we are sorry for this, this never should have happened. We will intensify our control procedure so that this does not happen again.
Accusation of immorality + extended apology		Our apologies, we are sorry for this, this never should have happened. This happened because of a software trouble in the logistics unit. We will intensify our control procedure so that this does not happen again.

Table 1. *Multilevel analyses of trustworthiness of the shops (eight evaluations per participant): Unstandardized coefficients followed by standard errors.*

Model	I	II	III	IV	V	VI
Trust disposition (H2)	0.22**	0.22*	0.21*	0.22**	0.21*	0.20*
	0.08	0.09	0.08	0.09	0.08	0.08
Age	0.06	0.09*	0.03	0.09*	0.03	0.04
	0.03	0.04	0.04	0.04	0.04	0.03
Easily spend money on gadgets	0.06*	0.07*	0.04	0.07*	0.04	0.03
	0.03	0.03	0.03	0.03	0.03	0.03
Reliability of Bol.com	0.22***	0.21***	0.23***	0.21***	0.23***	0.22***
	0.05	0.05	0.05	0.05	0.05	0.05
Morality accusation (H1)	-0.05	0.02	-0.11			
	0.04	0.06	0.06			
Broken product	-0.20*					-0.3
	0.1					0.15
Reputable store	1.16***	1.18***	1.14***	1.18***	1.15***	1.38***
	0.04	0.06	0.06	0.06	0.06	0.16
Comment is apology (H4)	0.39***	0.42***	0.69***			0.62***
	0.09	0.08	0.09			0.16
Comment is denial	-0.08	-0.05	-0.26**			-0.04
	0.09	0.08	0.09			0.17
Broken product X denials	-0.17					-0.05
	0.12					0.19
Broken product X apologies	0.32**					0.29
	0.12					0.19
Apology plain (H4)				0.43***	0.52***	
				0.09	0.11	
Apology extended (H4)				0.41***	0.84***	
				0.1	0.11	
Denial plain				-0.15	-0.37***	
				0.1	0.11	
Denial extended				0.04	-0.14	
				0.1	0.11	
Reputable store X denials (H6)						-0.12
						0.2
Reputable store X apologies						-0.39*
						0.19
Constant	1.59***	1.38**	1.58***	1.39**	1.53***	1.67***
	0.4	0.46	0.44	0.45	0.44	0.44
N	2312	1156	1156	1156	1156	1156
Cases	all cases	delayed delivery	broken product	delayed delivery	broken product	morality accusation

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 2

Multi-level analyses of the believability of the web shop's reactions (eight evaluations per participant).

Unstandardized coefficients followed by standard errors.

Models	VII	VIII	IX
Trust disposition	0.24**	0.26**	0.22*
	0.08	0.09	0.09
Age	0.02	-0.01	0.04
	0.03	0.04	0.04
Easily spend money on gadgets	0.06*	0.04	0.08*
	0.03	0.03	0.03
Reliability of Bol.com	0.21***	0.22***	0.20***
	0.05	0.05	0.06
Morality accusation	0		
	0.05		
Broken product	-0.02	-0.14	0.1
	0.05	0.08	0.08
Reputable store	0.68***	0.64***	0.59***
	0.05	0.11	0.12
Reaction is denial (vs apology) (H3)	-0.93***	-1.00***	-1.05***
	0.06	0.12	0.12
Reputable store X reaction is denial (H5)		0.1	0.17
		0.17	0.17
Constant	2.34***	2.52***	2.24***
	0.4	0.44	0.48
N	1859	942	917
Cases	only cases with reactions	morality accusations	competence accusations

Note. * $p < .05$, ** $p < .01$, *** $p < .001$