Essay 2:

The influence of consumer and expert ratings on the decision making for complex credence services

Abstract

Online retailers and specialized rating platforms often provide consumer ratings and quality assessments from experts as a decision aid for consumers. This study aims to assess the effects of those two different sources of online advice when provided simultaneously in the context of complex services. With three experimental studies and an analysis of clickstream data from a healthcare rating platform, we investigate how consumers make use of expert and consumer ratings, especially when these sources are presenting conflicting advice. We find that consumers are influenced more strongly by other consumers than by experts, but identify rating volume as an important moderator: only when consumer advice is based on a large number of underlying opinions, website users (i.e., consumers) follow it, otherwise they follow the expert rating. The stronger impact of consumer ratings is driven by the perceived similarity with the source of advice, which outweighs the lack of perceived expertise compared to expert sources.
Introduction

Online retailers and (independent) rating platforms provide website users (i.e., consumers) with different decision support tools such as ratings from other consumers, and supposedly more objective quality ratings based on expert opinions (hereafter referred to as ‘expert ratings’). Consumers use this information to reduce the uncertainty related to purchase decisions (Kim and Hollingshead 2015; Libai et al. 2010; Verlegh et al. 2013). While prior research has established the impact of both expert (e.g., Floyd et al. 2014) and consumer ratings (e.g., Dhar and Chang 2009) on consumer decision making (Kaptein and Eckles 2012), these studies have mainly studied these ratings in isolation, and have not assessed the impact of online expert and consumer ratings that are provided simultaneously (cf., Mudambi and Schuff 2010). As a result, little is known about how consumers consolidate those different sources of advice—which is especially relevant when those sources are conflicting. This question is not only of theoretical, but also of practical relevance, as several popular rating platforms provide expert and consumer ratings side by side (e.g., consumerreports.org, medicare.gov, nhs.uk, independer.nl).

The integration of expert and consumer ratings is particularly relevant in the domain of complex services (e.g., financial, legal, educational, or health services). Such services are typically considered high in credence qualities, i.e., they possess characteristics that consumers cannot properly evaluate even after consumption or use (Darby and Karni 1973), which makes it impossible for consumers “to evaluate whether these services are necessary or are performed properly” (Zeithaml 1981, p.186). Consequently, credence services often involve considerable levels of uncertainty, information asymmetry, and perceived risk (Mattila and Wirtz 2002; Ostrom and Iacobucci 1995: Pan and Chiou 2011). To reduce uncertainty, advice from other consumers might be less helpful, because those consumers have equal difficulty in
assessing the credence qualities, even after they have consumed or experienced the service. Experts, on the other hand, may be more likely to possess the skills and knowledge necessary to assess these credence qualities. In line with this notion, several legislators have implemented guidelines that require sellers to provide consumers with (independent) expert advice before purchasing or using complex credence services. In The Netherlands, for example, a recent change in Dutch legislation with regard to complex financial services (www.AFM.nl) requires that sellers provide consumers with expert advice before they can obtain such a service. Similarly, pharmacists in several countries are legally required to provide advice to patients before supplying any kind of medication (e.g., Aronson 2009; National Association of Boards of Pharmacy 2012). But in other cases, policy makers seem less concerned with the potential lack of expertise for consumers to make a judgment and use consumer ratings as a base for policy making. Think for example of the (planned) resource allocation policies based on consumer satisfaction in healthcare in both the UK and the US (e.g., BBC 2015; Medical Group Management Association 2014). Although there has been little research on the use of online ratings for complex credence services, we know from other contexts that consumers tend to heavily rely on consumer ratings (De Langhe, Fernbach, and Lichtenstein 2016) and often prefer advice from their peers over advice from experts (e.g., Dhar and Chang 2009; Huang and Chen 2006). Such a preference for consumer advice seems logical for search (e.g., computer) and experience goods or services (e.g., hotels) as consumers acquire knowledge about their relevant features while using or consuming them (Vermeulen and Seegers 2009), but has less relevance for credence goods and services. Hence, the aim of the current study is to explicitly assess how consumers make use of simultaneously provided online ratings from consumers and experts in a credence service context.
The current study contributes to theory and practice in three important ways. First, we investigate how consumers make decisions about complex credence services based on simultaneously provided consumer and expert ratings. While previous research primarily focused on the effects of either consumer or expert ratings on consumer decision making (e.g., Senecal and Nantel 2004), we focus on situations where both types of ratings are provided simultaneously. This design feature of, for instance, online rating platforms is not only of increasing practical interest. Importantly, it also allows for disagreement between the consumer and expert advice, which is of theoretical interest because it forces consumers to make a decision about which source they want to follow. Specifically, we test the—at first glance counterintuitive—idea that consumers base their decisions more strongly on advice from other consumers than from experts, although the latter may be seen as more competent to give advice in the respective situation.

As a second major contribution, we analyze the underlying process that causes consumers to be influenced more strongly by one source over the other. Based on social influence theory, we argue (and find) that perceived similarity of other consumers drives the stronger influence of this kind of advice and even overcompensates the negative effect of lower perceived expertise of other consumers as compared to experts.

Finally, as our third contribution, we investigate whether this stronger influence of consumer over expert ratings is conditional on the number of underlying consumer opinions that contributed to the aggregated consumer rating (hereafter referred to as ‘rating volume’). A common design practice of online rating platforms is to present only one expert rating on the platform, whereas the consumer rating typically is based on the aggregated opinion of multiple consumers. Whether or not the number of consumers that contributed to the aggregated rating is stated on the platform, however, varies between platforms. Again, based on social influence theory, we argue that the influence of...
consumer ratings is dependent on the number of underlying consumer opinions that contributed to the overall rating. This is not only an important insight for practitioners that have to decide about the design of such a platform, it moreover further helps us to understand the psychological processes underlying consumers’ preferences for advice from other consumers or experts. Methodologically, we combine multiple experimental studies with field data to provide high levels of both external and internal validity.

**Literature review and hypotheses**

**Effects of consumer and expert advice**

Advice from experts as well as consumers has been found to heavily impact both consumer perceptions and financial performance of a product or firm (e.g., Bansal and Voyer 2000; Chevalier and Mayzlin 2006; Gilly et al. 1998). As Table 1 illustrates, a number of studies compare their impacts but do so in isolation by confronting participants with either type.
<table>
<thead>
<tr>
<th>Article</th>
<th>Context</th>
<th>Study type</th>
<th>Dependent variable</th>
<th>Source manipulation (participants confronted with either consumer OR expert rating)</th>
<th>Relevant finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feldman and Spencer</td>
<td>Advice on physicians</td>
<td>Survey</td>
<td>Choice of source of advice</td>
<td>Consumer n.a. Expert n.a.</td>
<td>85% of participants turned to similar individuals when looking for advice on a physician; only 15% turned to individuals with medical knowledge. Average consumer ratings are better predictors of future sales than average ratings by music media publications. Purchase intention influenced more strongly by consumer than expert recommendation.</td>
</tr>
<tr>
<td>Dhar and Chang (2009)</td>
<td>Online conversations on music</td>
<td>Secondary data</td>
<td>Sales</td>
<td>Consumer ratings Expert ratings by music media publications</td>
<td>Average consumer ratings are better predictors of future sales than average ratings by music media publications. Purchase intention influenced more strongly by consumer than expert recommendation.</td>
</tr>
<tr>
<td>Zhang et al. (2010)</td>
<td>Online reviews on restaurants</td>
<td>Experiment</td>
<td>Intention to visit restaurant's webpage</td>
<td>Consumer reviews Editorial reviews</td>
<td>Editor reviews negatively affect consumers’ willingness to visit the webpage of a restaurant, while consumer reviews increase it.</td>
</tr>
<tr>
<td>Flanagan and Metzger</td>
<td>Online ratings on movies</td>
<td>Experiment</td>
<td>Credibility perceptions</td>
<td>Consumer ratings Critics' ratings</td>
<td>Credibility perceptions were significantly higher for ratings provided by experts than consumer ratings. Reviews have a significantly higher impact on sales when they come from experts than from other consumers.</td>
</tr>
<tr>
<td>Floyd et al. (2014)</td>
<td>Effect of online reviews on sales</td>
<td>Meta-analysis</td>
<td>Sales</td>
<td>Consumer reviews Expert reviews</td>
<td>No significant differences between influence of consumer and expert recommendations. Recommender system had strongest impact.</td>
</tr>
<tr>
<td>Study</td>
<td>Context</td>
<td>Method</td>
<td>Findings</td>
<td>Consumers</td>
<td>Experts</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>Smith, Menon, and Sivakumar (2005)</td>
<td>Online advice on restaurants</td>
<td>Experiment</td>
<td>Choice of source</td>
<td>Consumer reviews</td>
<td>Editorial reviews</td>
</tr>
<tr>
<td>Chakravarty, Liu, and Mazumdar (2010)</td>
<td>Online reviews on movies</td>
<td>Experiment</td>
<td>Evaluation of movies</td>
<td>Consumer reviews</td>
<td>Critics' reviews</td>
</tr>
<tr>
<td>De Maeyer and Estelami (2010)</td>
<td>Online advice on cellular service provider/phone</td>
<td>Experiment</td>
<td>Credibility of and intention to use rating site</td>
<td>Consumer reviews</td>
<td>Expert reviews</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Zhang et al. (2010) find that positive expert reviews even reduce consumers’ willingness to visit a restaurant’s online presence in order to get more information, whereas consumer reviews increase it. Similarly, in an early survey on the choice of family doctors, two-third of the people reported seeking advice from peers rather than those with medical expertise (Feldman and Spencer 1965). Yet other studies indicate instead that expert advice influences consumers’ decisions more strongly than other consumers’ advice (Flanagin and Metzger 2013; Floyd et al. 2014; Woodside and Davenport 1974), suggesting the importance of a more nuanced view of consumer versus expert reviews.

This nuanced view is represented by a number of studies arguing that the impact of expert versus consumer sources depends on factors like shopping goals (Smith, Menon, and Sivakumar 2005), product category (De Maeyer and Estelami 2011), or receivers’ experience (Chakravarty, Liu, and Mazumda 2010). Thus, consumers appear to favor consumer advice for hedonic shopping, services, and when they have limited experience. They prefer expert advice instead for utilitarian shopping, products, and when they are more experienced. However, none of these studies assessed the impact of expert and consumer sources when provided simultaneously and in a credence context.

Social influence

To understand (and make predictions about) how consumers use online expert versus consumer ratings in their decision making, it seems valuable to explore why consumers attach such great importance to advice by others. According to social influence literature people use the behavior of others around them as descriptive norms on how to behave (Cialdini and Goldstein 2004; Schultz et al. 2007). People are more likely to perform a certain behavior when knowing that others are likewise doing it. This imitation of peers is deeply rooted in human evolution (Henrich 2015). It can reduce search costs, the risk of failure (Munshi 2004), and avoid experimentation (DiMaggio and Powell
This so-called social proof heuristic has been found to impact a wide array of actions such as littering in public places (Cialdini, Reno, and Kallgren 1990), charity donations (Reingen 1982), committing suicide (Garland and Zigler 1993), reusing hotel towels (Goldstein, Cialdini, and Griskevicius 2008), as well as brand choices and purchase decisions (e.g., Bearden and Etzel 1982; Godinho de Matos, Ferreira, and Krackhardt 2014; Griskevicius et al. 2009). The rise of the Internet has made the behavior and opinion of others increasingly visible, leading to an amplified likelihood of consumers being influenced by others, especially by others’ positive judgments (Muchnik, Aral, and Taylor 2013). Social influence literature suggests that it is strongest if a piece of advice comes from people that are similar to the decision maker (e.g., consumers; Cialdini and Goldstein 2004).

**Similarity**

Social influence is heavily driven by the similarity receivers perceive between the people giving advice and themselves (e.g., Cialdini and Goldstein 2004; Festinger 1954). Source similarity, or homophily, describes the extent to which people resemble one another on certain attributes (Brown and Reingen 1987). Similar sources tend to be particularly influential, because they share similar needs and preferences with the receiver and thus deliver relevant information (Brock 1965; Brown, Brodetrick, and Lee 2007; Simons, Berkowitz, and Moyer 1970). Similarity also leads to greater attractiveness, trust, and understanding (Ruef, Aldrich, and Carter 2003), which facilitates communication (Price and Feick 1984). Therefore, information from similar others is more persuasive and has a stronger influence on decision making than information from dissimilar others (e.g., Brown and Reingen 1987; Simons, Berkowitz, and Moyer 1970). However, not only socially meaningful similarities can create a perceived bond between people, but also simply the fact that they share a similar experience or situation (Goldstein, Cialdini, and...
Griskevicius 2008; Tajfel 1978). For instance, Goldstein et al. (2008) illustrate that consumers are more sensitive to information about the behavior of other consumers when these consumers have had a similar experience. In the present credence context, the decision maker might realize that other consumers have been in the same situation of needing a certain service. Experts instead might be perceived as being less similar as they do not share the same situation and probably approach the assessment of a service in a more abstract and technical manner. Such an out-group focus (Tajfel 1987) might make consumers assume that the expert rating contains less relevant information with regard to the consumers’ individual and special needs. Taken together, we predict that consumers perceive other consumers to be more similar than experts, which in turn makes the advice of these other consumers more influential. We thus hypothesize:

**H1**: a) Consumer evaluations and b) usage intentions of a credence service are influenced more strongly by consumer than by expert ratings.

**H2**: Similarity perceptions drive the stronger influence of consumer versus expert ratings on a) consumer evaluations and b) usage intentions of a credence service.

The moderating role of rating volume information

Rating platforms very often not only provide star ratings but also inform users about how many individual consumer opinions have been aggregated to compute those ratings. However, this ancillary information is usually only given for consumer ratings, while expert ratings are not further qualified by specific volume information (e.g., consumerreports.org, nhs.uk, independer.nl). This design feature leads to the theoretically profound question whether information about rating volume (i.e., the amount of other consumers’ opinions that underlie a credence service rating) may constitute an important
moderator for the influence of such ratings on other consumers’ service evaluations and usage intentions.

Social influence literature states that the likelihood that an individual imitates a certain behavior increases with the number of other (similar) individuals that display this behavior (e.g., Cialdini and Goldstein 2004; Latané and Wolf 1981; Tanford and Penrod 1984). The high number of others increases the perceived benefits of adopting a certain behavior compared to its perceived costs (Lopez-Pintado and Watts 2008; Watts and Dodds 2007). Numerical information has been found to be automatically encoded by the human brain and thus easy to process relatively independent of individual differences (Hasher and Zacks 1984). The thus easily graspable numerical dominance of a majority opinion signals its correctness (Baker and Petty 1994; Latané 1981). Consumers have been found to be especially prone to the influence from a large number of others when it comes to making purchase decisions (e.g., Dholakia, Basuroy, and Soltysinski 2002; Liu 2006; Salganik, Dodds, and Watts 2006). Consumers engage in such imitative behavior because a high volume increases the diagnosticity of a message’s valence and thus its persuasiveness (Khare, Labrecque, and Asare 2011). This might be due to the fact that high consensus signals that those others might possess a certain piece of information that is not yet available to the decision maker, (i.e., "they must all know something I don't"; Dholakia, Basuroy, and Soltysinski 2002; Prendergast and Stole 1996). In line with that, the credibility as well as the impact of a group’s judgment on an individual have been found to be positively correlated with the group’s size—both online and offline (e.g., Duan, Gu, and Whinston 2008; Flanagin and Metzger 2013; Ye et al. 2011). Hence, rating volume is likely to amplify the effect of advice from similar consumers (compared to expert advice) on consumer evaluation and decision making. In other words: the larger the
number of consumers that rate a complex credence service, the more influential the (averaged) consumer rating. We thus hypothesize that:

**H3:** The influence of consumer (versus expert) ratings on a) consumer evaluations and b) usage intentions of a credence service is moderated by rating volume, such that with increasing numbers of underlying consumer opinions, the influence of the consumer over the expert rating increases.

**Study 1a**

With Study 1a, we assess how the simultaneous exposure to conflicting consumer and expert ratings affects website users’ evaluations of a service provider (H1) in a controlled experimental setting. Furthermore, we analyze the underlying process and especially test the mediating role of perceived similarity (H2). We used a healthcare context as an example for credence services. As patients usually lack medical knowledge and outcomes of procedures are often only fully realized months or even years after the service provision, healthcare services are especially high in credence qualities (Berry and Bendapudi 2007; Zeithaml 1981).

**Procedure and Sample**

We employed a one-factorial between-subjects design, in which we manipulated the valences of the expert rating (positive vs. negative) and consumer rating (positive vs. negative). We included only manipulations of conflicting ratings (i.e., expert positive and consumer negative and vice versa - we refer to this variable as ‘positive source type’, as it can be ‘consumer positive’ or ‘expert positive’). We recruited U.S. participants through Amazon’s Mechanical Turk, in exchange for payment, and received 110 completed online questionnaires from respondents with an average age of 39.8 years, of whom 54.5% were women. We then randomly assigned them to one of two experimental conditions.
Participants were asked to imagine that they were looking for a hospital to perform knee surgery and for that purpose, were going to consult an independent rating platform to help make their decision (see Appendix A). On the following page, a fictional healthcare rating platform featured the evaluation of a fictional hospital, including both an expert and a consumer star rating, which were either negative (1.5 of 5 stars) or positive (4.5 of 5 stars).

Next, participants indicated their attitudes toward the hospital (Becker-Olsen 2003; Rodgers 2003) and their usage intentions (Chandran and Morwitz 2005). These variables correlated ($r = .84$, $p < .01$), so we combined them into a single evaluative index (‘hospital evaluation’; $\alpha = .97$). We also measured participants’ evaluations of experts and consumers with regard to their expertise (experts: $\alpha = .94$; consumers: $\alpha = .90$; Ohanian 1990), attitudinal similarity (experts: $\alpha = .96$; consumers: $\alpha = .94$; McCroskey, Richmond, and Daly 1975), and trustworthiness (experts: $\alpha = .95$; consumers: $\alpha = .94$; Newell and Goldsmith 2001). All these measures featured five-point scales.

**Results**

As a manipulation check, we assessed whether participants perceived experts as more knowledgeable than consumers. Indeed, experts scored significantly higher on expertise ($M_{\text{expert}} = 3.91$, $SD = .96$; $M_{\text{consumer}} = 3.21$, $SD = 1.01$; $t(109) = −5.03$, $p < .001$). Consumers were perceived to be more similar ($M_{\text{expert}} = 2.67$, $SD = 1.10$; $M_{\text{consumer}} = 3.83$, $SD = .087$; $t(109) = 9.32$, $p < .001$) and trustworthy than experts ($M_{\text{expert}} = 3.31$, $SD = .97$; $M_{\text{consumer}} = 3.69$, $SD = .84$; $t(109) = −3.08$, $p = .003$).

To test H1, we applied an analysis of variance (ANOVA), with hospital evaluation (i.e., the combined evaluative index) as dependent variable. The results showed a main effect of the positive source type (consumer positive vs. expert positive; $F(1, 108) = 6.40$, $p = .013$). The hospital was evaluated more favorably when the consumer rating was
positive and the expert rating negative (M_{consumer positive/expert negative} = 2.74, SD = 1.08) than when the consumer rating was negative and the expert rating positive (M_{consumer negative/expert positive} = 2.24, SD = 1.02). Thus, we find support for H1. To test the mediating effect of perceived similarity (H2), we conducted a mediation analysis with perceived similarity and perceived expertise of both the positive and negative sources as competing mediators.\(^2\) The results indicated that the effect of the positive source type on the evaluation of the hospital is fully mediated through similarity and expertise perceptions of the positive source. The indirect effects through both mediators were significant (Figure 1), but stronger for similarity of the positive source (indirect effect = .50, bootstrap 95% confidence interval [CI]: [.24, .81]) than for expertise of the positive source (indirect effect = −.18, bootstrap 95% CI: [−.42, −.02]). We did not find significant indirect effects through the other two mediators (similarity negative source: indirect effect = −.08, bootstrap 95% CI: [−.35, .14]; expertise negative source: indirect effect = .02, bootstrap 95% CI: [−.11, .16]).

\(^2\) This was necessary to be able to estimate the mediation effects as the perceived expertise and similarity of the peer and expert sources are invariant to the experimental condition.
Figure 1 Study 1a: Mediation analysis

Discussion

Study 1a provides evidence for both H1 and H2 by directly manipulating the valence of expert and consumer ratings in a controlled experimental environment. Also, when provided simultaneously, website users seem to be influenced more strongly by consumer than by expert ratings. This effect is driven by the higher perceived similarity of consumers, which overcompensates the higher perceived expertise of experts. Interestingly, it seems to primarily be the similarity and expertise of the positive source that drives the overall effect, which was not expected. This finding is, however, in line with the asymmetric herding effects found by Muchnik and colleagues (2013). The authors showed that the positive effect of positive online ratings on subsequent ratings was greater than the negative effect of negative ratings, suggesting that consumers are more likely to be influenced by positive social influence. While positive ratings were found to only create positive opinion change, negative ratings also triggered countervailing opinion change.
A possible explanation for the finding that website users seem to follow the consumer evaluation rather than the expert evaluation might be that users assume that consumers, different to experts, focus in their evaluations on aspects that are of higher importance for them to evaluate the credence service. Experts may be perceived to focus on technical quality and more objective measures. Consumers, on the other hand, may be expected to base their ratings on their subjective experience. This effect may be particularly strong in a healthcare context where patients’ subjective experience can be crucial for their psychological health (e.g., in extended hospital stays) and in turn contributes to their physical health (e.g., Kolappa, Henderson, and Kishore 2013). In this context website users may thus deliberately attach more weight to consumer ratings. To assess whether these findings are generalizable to other contexts in which consumers’ subjective experience is less crucial than a technically sound and efficient execution, Study 1b replicates them for an auto repair service.

**Study 1b**

With Study 1b, we aim to replicate and extend our findings from Study 1a to a different context.

**Procedure and Sample**

We employed the same procedure as in Study 1a but using an auto repair context. We chose this context because, different to the healthcare context, in the auto repair context the subjective experience and psychological health of the consumer can be assumed to be unrelated to the service outcome (i.e., the quality of the auto repair). Thus, consumers can be expected to attach more weight to the expert rating than in the previous healthcare context. We asked participants to imagine that they consulted a rating platform to help make their decision when looking for an auto repair service after a car break-down. We again recruited U.S. participants through Amazon’s Mechanical Turk and received 107
completed online questionnaires from respondents with an average age of 39.3 years, of whom 46.7% were women. We then randomly assigned them to one of two experimental conditions.

After consulting a fictional rating platform that featured both an expert and a consumer star rating of a fictional auto repair service (see Appendix B), participants indicated their attitudes toward the auto repair service (Becker-Olsen 2003; Rodgers 2003) and their usage intentions (Chandran and Morwitz 2005). Again, these variables correlated ($r = .77$, $p < .01$), so we combined them into a single evaluative index (‘auto repair evaluation’; $\alpha = .95$). Analogous to Study 1a, we also measured participants’ evaluations of experts and consumers with regard to their expertise (experts: $\alpha = .95$; consumers: $\alpha = .89$; Ohanian 1990), attitudinal similarity (experts: $\alpha = .93$; consumers: $\alpha = .93$; McCroskey, Richmond, and Daly 1975), and trustworthiness (experts: $\alpha = .96$; consumers: $\alpha = .94$; Newell and Goldsmith 2001).

**Results**

Like in Study 1a, experts scored significantly higher on expertise than consumers ($M_{\text{expert}} = 3.91$, $SD = .89$; $M_{\text{consumer}} = 3.10$, $SD = .90$; $t(106) = -6.56$, $p < .001$). Consumers were perceived to be more similar ($M_{\text{expert}} = 2.63$, $SD = .88$; $M_{\text{consumer}} = 3.63$, $SD = .84$; $t(106) = 8.64$, $p < .001$) but equally trustworthy as experts ($M_{\text{expert}} = 3.35$, $SD = .87$; $M_{\text{consumer}} = 3.49$, $SD = .84$; $t(106) = -1.22$, $p = .224$).

To find further support for H1, we applied an ANOVA with auto repair evaluation as dependent variable. The results showed a main effect of the positive source type (consumer positive vs. expert positive; $F(1, 105) = 9.15$, $p = .003$). The auto repair service was evaluated more favorably when the consumer rating was positive and the expert rating negative ($M_{\text{consumer positive/expert negative}} = 2.76$, $SD = .91$) than when the consumer rating was negative and the expert rating positive ($M_{\text{consumer negative/expert positive}} = 2.27$, $SD = .75$). Thus,
we again found support for H1. To test the mediating effect of perceived similarity (H2), we conducted a mediation analysis with perceived similarity and perceived expertise of the positive and negative source as competing mediators. The results indicated that the effect of the positive source type on the evaluation of the auto repair service is fully mediated through similarity and expertise of the positive source. The indirect effects via both mediators were significant (Figure 2), and again stronger for similarity (indirect effect = .35, bootstrap 95% CI: [.15, .64]) than for expertise (indirect effect = −.24, bootstrap 95% CI: [−.45, −.10]). We did not find significant indirect effects through the other two mediators (similarity negative source: indirect effect = .03, bootstrap 95% CI: [−.12, .19]; expertise negative source: indirect effect = −.01, bootstrap 95% CI: [−.15, .11]).

**Figure 2** Study 1b: Mediation analysis

**p < .05. ***p < .001.**
Discussion

Study 1b was designed to test our main effect and underlying process in a context in which a consumers’ subjective experience during the service can be expected to play a less prominent role compared to the objective service outcome. Both Study 1a and 1b find support for our first two hypotheses. Across two different complex credence service contexts, website users are influenced more strongly by consumer than by expert ratings. When simultaneously confronted with those two sources of advice, perceived similarity plays a greater role than perceived expertise in determining which advice to adopt. Following the example or advice of similar others (social proof heuristic, see above) when making decisions on search or experience goods or services seems logical as their subjective experience might indeed be indicative of users’ own enjoyment. For complex credence services, however, this strategy might be less straightforward as there is an objective quality of the service delivery that should (at least) be equally important as the subjective experience.

Studies 1a and 1b presented participants with consumer and expert star ratings without any ancillary information which is crucial to isolate the differential effects of both sources of advice. However, increasingly rating platforms also feature the number of underlying consumer opinions that have been aggregated to form an overall star rating. It is important to also analyze the potential moderating role of this interesting design feature as (a) such volume measures have been found to be highly influential when assessing the impact of consumer ratings (e.g., Khare, Labrecque, and Asare 2011) and (b) they are usually only provided for consumer, but not for expert ratings.

Study 2

Study 2 relies on clickstream data from an actual U.S. healthcare rating platform to assess the moderating role of explicit rating volume information for the dynamic discussed
above. Additionally, Study 2 also replicates the findings from the two experimental studies in a real-life setting and thus contributes to the external validity of the current paper.

The U.S.-based rating platform provides expert and consumer ratings on more than 1 million doctors, dentists, and eye doctors throughout the United States to help consumers find a suitable service provider. For each doctor, an algorithm calculates an expert rating, on the basis of the physician’s education, experience, training, and referrals by other doctors. It also displays the average Yelp (consumer) rating for each doctor. Both rating types are presented next to each other on overview and search results pages from which consumers can access the individual doctor profiles. We assessed the impact of both rating types on two behavioral actions: the number of times website users viewed a certain doctor’s profile and the number of times they clicked a button to obtain the doctor’s contact information.

**Data Collection**

We obtained clickstream data for 5,299 doctors during May 1–July 31, 2015 from the rating platform’s web analytics tool and aggregated the click-level data (21,897 profile clicks and 1,842 call clicks, see below) to the doctor-level (n = 5,299). We only included doctors whose profiles featured both expert and consumer ratings, and whose profiles had been viewed at least once by a prospective consumer during the data collection period. We extracted how many times each doctor’s profile had been viewed and how many times users clicked to obtain the doctor’s contact information. Furthermore, for each doctor we extracted the expert star rating, the Yelp (consumer) star rating (both on five-point scales in .5 steps, with 5 = best), and the rating volume the overall Yelp rating was based on at the moment of the click. Additionally, we collected several control variables (see below).

**Measures**
We have two dependent variables: number of profile clicks and number of call clicks. That is, we count the number of times each doctor profile was viewed by prospective consumers in the data collection period. We also determined how many times visitors to a specific doctor’s profile clicked on a button to obtain this doctor’s phone number. This behavioral measure provides a proxy for doctor choice.

The independent variables reflected expert and consumer ratings. Expert ratings were the professional star ratings provided for each doctor, which ranged from 1 to 5 in .5 steps (half stars). The average expert rating was very positive (M = 4.26, SD = .94). Consumer ratings were provided in the form of Yelp star ratings on each doctor’s profile. The Yelp ratings could change during the data collection period, we therefore used the respective Yelp rating at the moment of each individual click to calculate an average consumer rating for each doctor over the data collection period. Consumer ratings also varied from 1 to 5, in .5 steps, and the average consumer rating in our sample was positive too (M = 3.89, SD = 1.21). To be able to test the effect of rating volume, we included the number of individual consumer opinions the Yelp star rating was based on in our model. This number could also change over the course of the data collection period, so we again calculated an average rating volume for each doctor over the data collection period.

Finally, to isolate the effects of expert and consumer ratings on the click-based dependent variables, we included doctor-specific control variables. We controlled for the number of doctor referrals. Based on 160 million Medicare referrals from 2009 to 2012, this number indicates how many patients have been referred by other doctors to a specific doctor. Doctor referrals only appear on a profile when they are greater than 0, so we included a dummy variable for whether doctors had any referrals or not (0 = no, 1 = yes). Further, we controlled for each doctor’s specialty; the number of practices in which she or he is employed; and whether the doctor has a profile image (0 = no, 1 = yes), photo gallery.
(0 = no, 1 = yes), or online booking system incorporated in the profile (0 = no, 1 = yes), as well as whether a doctor has a premium profile (0 = no, 1 = yes) on the platform.³

**Specification**

Most doctors in our sample received few profile or call clicks during the data collection period, and a few doctors had many visitors. Thus, both dependent count variables are overdispersed (Mprofile clicks = 4.12, variance = 85.89; Mcall clicks = .34, variance = 3.03), and we therefore modeled both dependent variables with a negative binomial regression (Greene 2003). To account for systematic differences between different kinds of doctors, we included a random intercept for each of the 57 doctor specialties:

\[
\log(\text{profile clicks}_i) = \alpha_0 + \alpha_j + \beta_1(\text{expert rating}_i) + \beta_2(\text{consumer rating}_i) + \beta_3(\text{consumer rating}_i \times \text{rating volume}_i) + \Omega X_i + \varepsilon_i,
\]

where i indexes the doctor and j the doctor specialty, X_i is the vector of doctor-specific control variables, and \(\varepsilon_i\) is the error. The model specification for our second dependent measure, call clicks, is analogous.

**Results**

Table 2 shows the results of the regression models. The analyses for the profile clicks (call clicks) model exclude 5 (1,213) doctors due to missing values for the predictors, resulting in a sample size of 5,294 (4,086). In our two initial models (Models 1 and 3), we found a significant positive effect of the expert rating on both profile clicks (\(\beta = .13, p < .001\)) and call clicks (\(\beta = .32, p < .001\)) and of the consumer rating on profile clicks (\(\beta = .03, p = .022\)) but not on call clicks (\(\beta = .04, p = .761\)). The interactions of

---

³ For our call clicks model we controlled for all variables; for the profile clicks model we only controlled for doctor specialty, practice count, profile image, and premium profile as the other control variables only become visible to consumers once they have clicked on a doctor’s profile (and not on the search results page).
consumer rating and rating volume (mean centered; Jaccard and Turrisi 2003) on profile clicks ($\beta = .01, p < .001$) and call clicks ($\beta = .01, p < .001$) were, however, significant. To investigate these interaction effects further and assess the effect of the consumer rating on the dependent variables with different numbers of underlying individual ratings, we repeated the analyses with a mean-centered value for the rating volume plus three standard deviations (Models 2 and 4). As expected, the effect of the consumer rating increased with the number of individual reviews aggregated to create the rating (Figure 3). As we noted previously, the effects of consumer ratings were only significant for profile but not for call clicks for an average rating volume (Yelp rating based on 9 opinions), but their effects were significant and comparable to those of expert ratings, when the underlying number of consumer opinions was one standard deviation higher (Yelp rating based on 27 opinions; profile clicks $\beta = .14, p < .001$; call clicks $\beta = .28, p = .030$). The effects of expert and consumer ratings were not significantly different from each other: profile clicks $\beta = .01, \text{SE} = .04, p = .373$; call clicks $\beta = -.05, \text{SE} = .15, p = .380$. When the consumer rating was based on the average rating volume plus three standard deviations (Yelp rating based on 63 opinions), the effects on both dependent variables (profile clicks $\beta = .35, p < .001$; call clicks $\beta = .75, p = .015$) were even stronger. With the number of underlying consumer opinions increasing to 63, the effect of the consumer rating becomes (marginally) significantly stronger than the effect of the expert rating (profile clicks $\beta = .23, \text{SE} = .07, p = .004$; call clicks $\beta = .43 \text{SE} = .32, p = .100$).
<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>DV: log profile clicks</th>
<th>DV: log call clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Expert rating</td>
<td>Coeff.</td>
<td>SE</td>
</tr>
<tr>
<td>.13***</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Consumer rating × rating volume (centered at mean)</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Consumer rating (at mean level of rating volume)</td>
<td>.03**</td>
<td>.01</td>
</tr>
<tr>
<td>Rating volume</td>
<td>−.00</td>
<td>.00</td>
</tr>
<tr>
<td>Consumer rating × rating volume (centered at mean + 3SD)</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Consumer rating (at mean level of rating volume + 3SD)</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Rating volume</td>
<td>−.00</td>
<td>.00</td>
</tr>
<tr>
<td>Controls</td>
<td>Intercept</td>
<td>.72***</td>
</tr>
<tr>
<td></td>
<td>Log-likelihood</td>
<td>−12910.8</td>
</tr>
<tr>
<td></td>
<td>Akaike information criterion</td>
<td>−12910.8</td>
</tr>
</tbody>
</table>

*p < .1, **p < .05, ***p < .01.

Table 2 Study 2: Regression results
Figure 3 Study 2: Effect of expert and consumer ratings on (1) profile clicks and (2) call clicks, conditional on rating volume (i.e., number of underlying consumer opinions)

(1)

(2)
Discussion

The results of the analysis of field data thus indeed confirm rating volume (i.e., the number of underlying consumer reviews) to be an important moderator of the impact of consumer versus expert advice. Specifically, we find that when confronted with information from both sources simultaneously, website users tend to base their evaluations of complex credence services on expert rather than consumer advice in case the consumer advice is based on small numbers. However, when the group of consumers substantially grows in size, website users start to rely on them rather than the expert in making a decision, which supports Hypothesis 3. To sum up, users are indeed influenced more strongly by advice from other consumers than by advice from experts, but only when the consumer advice is based on a large number of individual opinions. This finding is in line with prior research stating that volume increases the perceived diagnosticity and the effect of word-of-mouth messages (e.g., Flanagin and Metzger 2013; Khare, Labrecque, and Asare 2011). To further validate these findings specifically for conflicting consumer versus expert advice in a controlled setting, Study 3 aims at finding further support for the influence of the number of underlying consumer opinions (H3) by replicating the findings from Study 2 in an experimental setting.

Study 3

Procedure and Sample

We employed a 2×2 between-subjects design, in which we manipulated the positive source type (consumer positive vs. expert positive) and the number of individual opinions the consumer rating is based on (high volume vs. low volume). Our objective was to compare a very high and a very low rating volume to be able to effectively contrast the respective effects. Drawing from prior research (Liu 2006), observations of a Dutch healthcare rating platform (independer.nl), and especially from our Study 2 dataset, we set
the high rating volume to 142 (among top 1% of number of underlying opinions in our dataset). We set the low rating volume to 3 so that the aggregated consumer rating resembled more than just a single but still only very few opinions. We recruited 125 undergraduate students from a Dutch university in exchange for course credit and randomly assigned them to one of four experimental conditions in a computer lab. We excluded 13 participants who did not recall that there was a consumer and expert star rating on the site. In our sample (n=112), participants had a mean age of 23.1 years and 60.7% were women.

Analogous to Study 1a, participants imagined that they were looking for a hospital to perform a surgery, so they consulted a rating platform to help make their decision (see Appendix C). On the following page, the evaluation of a fictional hospital, on a fictional healthcare rating platform, featured both an expert and a consumer star rating. Those ratings were conflicting and either negative (1.5 of 5 stars) or positive (4.5 of 5 stars). The consumer rating was either based on 3 or 142 individual opinions. Next, participants indicated their attitudes toward the hospital (Becker-Olsen 2003; Rodgers 2003) and their usage intentions (Chandran and Morwitz 2005) which we again combined into a single evaluative index (‘hospital evaluation’; α = .96). We also measured participants’ evaluations of experts and consumers with regard to their expertise (experts: α = .94; consumers: α = .83; Ohanian 1990) and trustworthiness (experts: α = .95; consumers: α = .88; Newell and Goldsmith 2001). All these measures featured five-point scales.

Results

The results of an ANOVA indicated no significant main effects of positive source type (consumer positive vs. expert positive) and rating volume (142 vs. 3) on hospital evaluation ($F_{positive source type}(1, 108) = .17, p = .678; F_{volume}(1, 108) = .52, p = .472$). However, as expected, we found a significant interaction effect of these two variables on
hospital evaluation ($F(1, 108) = 13.06, p < .000$). Further analyses demonstrated that when consumer ratings are aggregated from a high volume of underlying opinions, consumers’ evaluations are affected more strongly by consumer than expert advice ($M_{\text{consumer positive/expert negative}} = 3.06, SD = .94$; $M_{\text{expert positive/consumer negative}} = 2.55, SD = .89$; $F(1, 108) = 4.93, p = .028$). Conversely, when consumer ratings are aggregated from a low volume, participants are affected more strongly by expert compared to consumer advice ($M_{\text{consumer positive/expert negative}} = 2.36, SD = .76$; $M_{\text{expert positive/consumer negative}} = 3.01, SD = .81$; $F(1, 108) = 8.42, p = .004$; Figure 4). Thus, Study 3 finds further support for H3.

Additionally, we also investigated whether expert and consumer ratings in this experiment were perceived differently with regard to expertise and trustworthiness based on the additional rating volume information. Experts were generally perceived as higher in expertise ($M = 3.95, SD = .73$) than consumers ($M = 2.57, SD = .69$; $F(1,110) = 190.33, p < .001$), this however was independent of the number of underlying consumer opinions in the experimental condition ($p > .85$). Trustworthiness did not differ between experts and consumers, again independent of the number of underlying consumer opinions (all $p$-values > .20).
Discussion

The results of Study 3 further support the finding from Study 2 that website users are influenced more strongly by other consumers, but only when the consumer evaluation is based on a larger number of individual opinions. When the consumer rating has been aggregated from only a small number of individual opinions, website users are instead more inclined to follow the expert advice. However, we find that participants still perceive experts to be higher in expertise, even than a large number of other consumers. Moreover, website users do not seem to trust experts less than consumers, or assume any biases or ulterior motives. Consequently, users seem to be aware of the fact that also a large number of other consumers still lacks the expertise to judge the objective outcome quality of a
credence service and also do not trust them more than an expert giving a rating. Yet, they are more inclined to follow their peers’ compared to an expert’s advice when choosing a complex credence service.

**General discussion**

**Conclusion**

When simultaneously exposed to ratings on an online platform, users seem to generally favor consumer over expert advice when choosing a complex credence service. Thus, they seem more affected by a source that likewise lacks the skills and knowledge to evaluate the objective outcome quality of a credence service. We find that perceived similarity of the two sources drives the stronger influence of consumer over expert advice, and overcompensates the effect of the other consumers’ perceived lack of expertise. In line with social influence literature (Cialdini and Goldstein 2004; Latané and Wolf 1981; Tanford and Penrod 1984), we further identify rating volume as an important moderator of this effect: only when based on a moderate to large number of individual opinions, website users are more inclined to follow other consumers’ advice, while they follow the expert advice instead when the consumer rating is only based on a small number of other consumers. Our findings from Studies 1a and 1b thus suggest that consumers seem to simply assume that aggregate consumer star ratings are based on a large number of opinions when no explicit information on rating volume is provided.

**Theoretical implications**

To the best of our knowledge, this study is the first to analyze the impact of the **simultaneous** presence of conflicting ratings from consumer and expert sources on consumers’ decision making for credence services. Research that provides participants with one source offers conflicting evidence for the preference for consumer over expert advice (e.g., Flanagin and Metzger 2013; Zhang et al. 2010), but such studies cannot
explicate the disaggregated effects of advice from different sources on a receiver or, in other words, how consumers decide “which electronic word-of-mouth messages to adopt and which ones to reject” (King, Racherla, and Bush 2014, p. 176). In three experimental studies and a field study, we shed light on the differential effects of consumer and expert advice in a complex credence service context. Accordingly, we offer three main implications for interactive marketing and decision making research.

First, this study enhances the understanding of website users’ use of decision support tools in the context of complex credence services. More specifically, we shed light on the integration of simultaneously provided information from other consumers and experts. We find that other consumers strongly influence users’ evaluations and choices of service providers, even overruling the contrasting opinions of experts. In this sense, choosing a provider of a complex credence service does not seem to differ much from purchase choices for (for example) movies or restaurants. Even for credence services like healthcare, others’ subjective ratings of service experiences have a greater impact on decision making than more objective ratings of the service. Although the subjective customer experience is also important in the context of complex credence services, it often fails to acknowledge the actual outcome quality of the service which consumers have difficulties evaluating even after it has been performed (Darby and Karni 1973). These two aspects should be considered complimentary for well-informed decisions, but our study suggests that this combination is not the route most consumers take when considering different sources of quality information on an online rating platform. Even in healthcare contexts, users turn to consumer over expert advice, and this negligence of objective outcome quality information might lead them to make suboptimal choices, with—in the case of healthcare—considerable consequences for their health and society at large.
Second, we contribute to literature on source effects by establishing the processes underlying users’ preference for consumer over expert advice even in situations in which other consumers’ advice might not be a suitable predictor of outcome quality. We show that the stronger influence of consumer advice in a complex credence context is counterintuitively driven substantially more so by perceived levels of similarity than perceived levels of expertise of the two sources of advice. In line with Kang and Herr (2006), we posit that this result might signal that consumers rely on heuristic processing (i.e., the social proof heuristic) to make complex, uncertain decisions in credence service contexts like healthcare (Berry and Bendapudi 2007).

Third, we identify consumer rating volume as an important moderator of this effect in that it can reverse the preference for consumer advice and make users follow the expert in case that the consumer rating is only based on a small number of individual opinions. By doing so, we contribute to the emerging literature that provides a more nuanced view of consumers’ use of different sources of advice by explaining seemingly contradictory findings from prior research with moderating effects (Chakravarty, Liu, and Mazumda 2010; De Maeyer and Estelami 2011; Smith, Menon, and Sivakumar 2005). Specifically, when no explicit volume information is present, consumers (in our studies) seem to assume a large number of aggregated opinions.

**Managerial implications**

Our findings provide helpful managerial insights for online retailers, providers of rating platforms, and providers of credence services. These insights were derived (in part) in discussions with credence service (i.e., healthcare) managers, doctors, and consultants.

For online retailers and platform providers, our results support the trend of providing advice from both expert and consumer sources. Experts and consumers differ in their expertise and access to information, so they often emphasize different features in their
judgments. These two sources of advice therefore should be regarded as complementary instead of substitutive input. Furthermore, the platform providers need to include the number of individual opinions underlying a consumer rating, because consumers might assume a rating is based on many opinions if the information is not evident. When this information is present, it acts as an important moderator, preventing an overemphasis on a collection of just a few consumer ratings. Similarly, by explicitly separating service experience and outcome quality judgments, a platform could prompt consumers to better evaluate the accessible information. Compared to overall aggregated ratings that seem to be processed in a heuristic way, these specific ratings might enable consumers to make explicit choices on what factors are more important to their respective situation. In general, platform providers need to recognize the potential detriments of oversimplifying the complexity and multifaceted nature of complex credence services such as healthcare, education, or financial advice.

For providers of credence services, our results imply that they should be acutely aware of the fact that consumers are influenced in their decision making by other consumers’ experiences, and perhaps more so than is commonly accepted by the industry (Sonnenberg 2014). A pure focus on outcome quality, and neglecting consumers’ service experiences, can have negative effects on client numbers. Even if they are evaluated favorably by experts, service providers need to deliver positive experiences and encourage consumers to rate them online. In contrast, negative expert ratings can be compensated for by a large number of satisfied clients.

Limitations and further research

In providing initial insights into the use of conflicting advice from different sources in decision making for credence services, this study features several limitations that provide directions for further research. First, in our experimental studies we did not specify
who the ‘experts’ were. Prior research indicates that users mainly look for authority cues when assessing a website’s credibility, but they rarely investigate who the expert sources actually are (Eysenbach and Köhler 2002). The participants in our studies perceived the experts as such. Still, it would be interesting to explore whether the findings change when the advice comes from various experts, such as family doctors or specialist physicians in our researched healthcare context.

Second, we did not analyze the potential moderating roles of other rating and platform characteristics. Platforms tend to feature both numerical ratings and textual reviews. Other important characteristics, including language use, salience of the valence, or further information about the sources and their reputation (King, Racherla, and Bush 2014) thus deserve further inquiry.

Third, this study investigated online ratings, one of many potential information sources a consumer employs to make credence service decisions. Offline sources such as friends and family also strongly influence credence service choice (e.g., Hoerger and Howard 1995). Further research might examine the differential effects of offline versus online word-of-mouth and their interaction, across the different stages of the decision making process. In turn, researchers could scrutinize the impact of rating platforms on a broader set of informational cues.
References


McCroskey, J. C., Richmond, V. P., and Daly, J. A. (1975). The development of a measure of perceived homophily in interpersonal communication. *Human Communication Research, 1*, 323-332.


