Electronic Commerce Research and Applications 28 (2018) 102–113

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

How do sellers use live chat to influence consumer purchase decision in China?

Zhepeng Lv, Yue Jin, Jinghua Huang *

Research Center for Contemporary Management, School of Economics and Management, Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 17 September 2017
Received in revised form 11 January 2018
Accepted 11 January 2018
Available online 31 January 2018

Keywords:
Behavioral research
Big data
China
Communication
Data analytics
E-commerce
Live chat
Purchase decisions
Social presence
Uncertainty reduction

ABSTRACT

Uncertainty brought about by the separation of information flow and product flow has become a critical obstacle to e-commerce development. From the perspective of presence and uncertainty, we attempt to determine whether live chat usage can influence consumer purchase decision and how live chat can be used to do so. Logit regression models are adopted to analyze data collected from an online store in Taobao.com. We find that: (1) live chat usage is positively associated with consumer purchase decisions; and (2) the behavior of sellers when using live chat can affect consumer purchase decision.

1. Introduction

With the development of the Internet, online shopping has become increasingly important to consumers. The retail sales of physical goods in China's online market reached 4190 billion RMB in 2016, accounting for 12.6% of the total retail sales of consumer goods with a year-on-year growth of 25.6% (Ministry of Commerce of the People's Republic of China, 2017). However, consumers cannot experience products (e.g., by trying on shoes) when they shop online. This limitation reduces presence and increases uncertainty, which in turn hampers purchase decision.

Certain techniques are available to assist consumers in their online shopping. Online product descriptions can provide basic information, and customer reviews can convey previous consumers' opinions. In addition, some shopping sites have launched live chat services to facilitate communication between consumers and sellers. For example, the well-known Chinese e-commerce site Taobao.com provides the frequently-used live chat "Ali Wangwang." Morgan Stanley reported that buyers and sellers may clarify their desires and demand through Ali Wangwang to enhance online transactions (Ji and Meeker, 2005). However, other shopping sites (e.g., Amazon) do not use live chat. eBay, another shopping site, even sold Skype in 2009, which was acquired for $2.6 billion in 2005 (Musil, 2009). Our study asks whether using live chat affects consumer purchase decisions. If so, how to use live chat for influencing consumer purchase decision must be addressed.

Several studies examine the relationship between live chat usage and consumer purchase decision. Some studies focus on the effect of live chat usage on presence in the conversation, which positively affects the trustworthiness of the relationship between consumers and sellers. Such a favorable effect facilitates the formation of consumers' purchase intention (Lu et al., 2016) and purchase decision (Ou et al., 2014). Jiang et al. (2010) argued that live chat usage facilitates the perceived interactivity (a key aspect of presence) of consumers, leading to their affective involvement and further purchase intention. Kang et al. (2014) emphasized the role of live chat usage on uncertainty reduction by increasing perceived interactivity. Tan et al. (2016) reported the positive effect of live chat usage on consumer purchase decision for low reputation sellers, arguing that live chat usage can help consumers reduce product quality uncertainty. The above literature confirms that live chat usage affects purchase decision by increasing presence. However, these studies fail to address how to use live chat

* Corresponding author.
E-mail addresses: lvzhp.09@sem.tsinghua.edu.cn (Z. Lv), Jiny.09@sem.tsinghua.edu.cn (Y. Jin), huangjh@sem.tsinghua.edu.cn (J. Huang).

https://doi.org/10.1016/j.elerap.2018.01.003
1567-4223/

to affect presence. In the information searching stage of users' decision-making process, the behavior of information providers (e.g., speed of response, frequency, and type of information) is relevant to the value of information due to their effect on presence (Song and Zinkhan, 2008) and uncertainty reduction (Weiss et al., 2008). In the context of live chat usage, our study attempts to analyze the effect of seller behavior as information providers on social presence and uncertainty reduction that influence purchase decisions.

We take steps to solve these issues and fill the research gaps. Our research: (1) re-examines the effect of live chat between consumers and sellers on consumer purchase decision; and (2) explores the behavior (response time, number of replies, and type of information) of sellers in using live chat to increase presence and reduce uncertainty. This study provides additional evidence regarding the role of live chat in the online shopping context. It also underlines the importance of the approaches in using live chat on a seller perspective. This exploration offers comprehensive insights into the antecedents of online purchase decision and the application of Burton-Jones and Straub's (2006) systematic conceptualization of information system (IS) usage in online shopping.

The rest of this study is organized as follows. In Section 2, we introduce related work on the effect of seller behavior on live chat usage. In Section 3, we present our hypotheses and arguments to justify them. In Section 4, we describe our data and specify the model. In Sections 5 and 6, we present our results and conclude with a discussion of the study's contributions and implications for theory and practice.

2. Literature review

2.1. Online shopping uncertainty

Online shopping uncertainty is the degree to which consumers fail to make predictions regarding a product or a firm they are dealing with (Pavlou et al., 2007). This uncertainty includes those related to sellers and products. Many scholars believe that *seller uncertainty* is mitigated by trust-building mechanisms, such as feedback ratings and third-party escrows (Dimoka et al., 2012; Benbasat et al., 2008). *Product uncertainty* is divided into product quality and product fit uncertainties. *Product quality uncertainty* refers to the difficulty to evaluate product attributes and predict future product performance (Ghose, 2009). *Product fit uncertainty* is the degree to which consumers cannot assess whether the product attributes match their preference (Hong and Pavlou, 2014). The problem of product fit uncertainty becomes prominent because consumers cannot test products when they shop online.

Prior studies explore the antecedents of product uncertainty. A majority of the studies confirm presence and product information as crucial factors for reducing product uncertainty (Weathers et al., 2007; Dimoka et al., 2012; Kang et al., 2014). Presence is discussed in the next section, given its considerable role in reducing product uncertainty and influencing other important variables (e.g., developing trustworthy relationship). Product information may come from sellers (e.g., product webpage descriptions and third parties). Product webpage descriptions (e.g., textual product description and product pictures provided by sellers) can convey comprehensive and vivid information, thereby increasing the confidence of consumers to form product evaluation (Peck and Childers, 2003). Third-party information (e.g., online reviews and consumer reports) can also validate postings of sellers regarding their products. Moreover, third-party information can provide additional cues about product attributes for consumers to evaluate products. Other studies show that information providers' behavior, such as timeliness and frequent response of other consumers in product forum (Weiss et al., 2008; Adjei et al., 2010), positively affects product uncertainty reduction. These factors are discussed later.

Product uncertainty reduction may drive consumers toward a purchase decision. Pavlou et al. (2007) stated that product uncertainty may result in suffering a loss. Consequently, consumers reduce purchase intentions and prevent purchase decisions. Adjei et al. (2010) found that consumers in online brand communities purchase products after product uncertainty is reduced.

Studies regarding product uncertainty classification and antecedents provide a basis for us to explore and analyze whether and how live chat can be used to reduce product uncertainty, thereby affecting the purchase decision of consumers. In addition, product webpage descriptions, online reviews, and ratings are examined in this study with reference to the above literature.

2.2. Presence of online shopping

Presence is generally defined as the perception of intimacy or being close to another person (Lowry et al., 2009; Short et al., 1976), including two dimensions, namely, *social presence* (the feeling of psychological closeness and interpersonal interaction in a mediated communication) and *telepresence* (the feeling of being physically engaged in a mediated environment). Regarding social presence, Lu et al. (2016) proposed that social presence in the web, perception of other consumers, and social presence of interaction with sellers are three important aspects of social presence in the online marketplace. Social presence can directly reduce product uncertainty or indirectly through trust toward sellers. Telepresence theory holds that perceived interactivity is a considerable aspect of telepresence. Perceived interactivity can attenuate the negative effect of product uncertainty by enhancing the ability of consumers to obtain information (Weathers et al., 2007; Kang et al., 2014). This feature can also help consumers study the attitudes and integrity of sellers, which stimulates the affective involvement of consumers by making them feel warm and pleased and creates a positive effect on the loyalty and attitude of consumers (Jiang et al., 2010; Song and Zinkhan, 2008). In this study, presence is used to evaluate and analyze the behavior of sellers in live chat usage to reduce product uncertainty, which further leads to purchase decisions.

2.3. Measurement of information system usage

IS usage is a classic research topic in the IS field (DeLone and McLean, 1992; Burton-Jones and Straub, 2006). Although researchers occasionally adopt specific measures for IS usage, they generally deploy similar measures (e.g., use or nonuse, duration of use, heavy or light use, frequency of use, and number of features used) (Alavi and Henderson, 1981; Venkatesh and Davis, 2000; Saga and Zmud, 1994). Burton-Jones and Straub (2006) proposed a systematic approach to reconceptualize IS usage by combining the user, system, and task; moreover, they suggested individual-level system usage as "an individual user's employment of one or more features of a system to perform a task."

The above literature provides a basis to measure live chat usage in an online shopping context. Live chat is a kind of IS for sellers and consumers to communicate with each other. Sellers' usage of live chat can affect consumers. In live chat, sellers are tasked to satisfy consumer needs, such as shortening the perceived distance between consumers and sellers online by responding rapidly and reducing product uncertainty by providing relevant and adequate information. Therefore, the way sellers use live chat or their behavior (e.g., frequency of use) can represent measurement of live chat usage in the online shopping context.

2.4. Behavior of information providers

In the information searching stage of users' decision-making processes, the behavior of information providers is relevant to the value of information. Weiss et al. (2008) conceptualized the following behavior of information providers relevant to information value: timeliness, frequency, and amount of information. *Timeliness* refers to the speed of responses to queries of the seeker. *Frequency* pertains to the number of contacts between information seekers and information provider. *Amount of information* denotes the total quantity of information provided to the seeker. Weiss et al. (2008) stressed the role of information providers' behavior to reduce information ambiguity and remove uncertainty. In addition, Song and Zinkhan (2008) proposed the effect of quick response on presence and underlined the required relevance of information type vis-à-vis the queries to influence presence and satisfaction.

As information providers, the behavior of sellers in live chat usage plays a consequently essential role in information value and subsequent decisions of consumers. By referring to the behavior of information providers, this study attempts to explore the effect of seller behavior in live chat usage on consumer purchase decision.

2.5. Limitations of previous studies

There has not been much literature that has discussed the role of live chat in the online shopping context (Jiang et al., 2010; Ou et al., 2014; Kang et al., 2014; Lu et al., 2016; Tan et al., 2016). Only two studies used actual transaction data to determine the relationship between live chat usage and consumer purchase decision. However, live chat is crucial in conveying information to consumers on products and helping establish connections between consumers and sellers. Consequently, there is an urgent need or researchers to provide additional evidence to address the increasing importance of live chat usage for consumer purchase decisions.

All the studies have focused on how live chat influences consumer purchase decision-making through presence and uncertainty. However, none of them discovered how to use live chat to influence presence and uncertainty, and subsequently contribute to consumer purchase decisions. According to the prior literature, the behavior of information providers can affect presence and uncertainty in communication (Weiss et al., 2008; Song and Zinkhan, 2008). IS usage studies also have suggested that the extent of use will make a difference in task performance (Burton-Jones and Straub, 2006). Hence, exploring behavior of sellers in live chat usage is necessary to determine how it should be used, which in turn will help to guide consumers toward making purchase decisions.

Some studies have found that information relevant to queries of consumers can increase presence (Song and Zinkhan, 2008) and reduce uncertainty (Adjei et al., 2010). In live chat usage, sellers frequently provide relevant information to queries of consumers. However, of all the relevant information, no literature has discussed the classification and effect of each type. This study aims to fill this gap by classifying relevant information into product quality-related and fit-related contents based on the uncertainty reduction literature. We further address the effect of each type on consumer purchase decisions.

3. Research hypotheses

3.1. Effects of live chat on consumer purchase decisions

Live chat can help consumers reduce product uncertainty by providing product information and increasing presence through communication with sellers. For product information, consumers (as information seekers) acquire significant information by communicating with sellers through live chat, thereby reducing a large set of possible beliefs regarding products (e.g., quality and fit issues), to a small set. Hence, uncertainty is eventually reduced.

For the effect on social presence, on the one hand, symbols (e.g., smileys and videos) affect the development of intimacy perception (Ou et al., 2014). The effective use of live chat (e.g., using humorous language, pictures, or emoticons) can facilitate mutual awareness between consumers and sellers, making the conversation similar to traditional face-to-face communication, which in turn increases social presence. On the other hand, live chat enables an interactive channel (Jiang et al., 2010), which helps accelerate verification and negotiation processes, thereby enhancing interactivity. The increased social presence and perceived interactivity mean a high level of presence. The high level of presence and increased trust because of presence (Lu et al., 2016) together contribute to removing product uncertainty (Pavlou et al., 2007). The level of product uncertainty is negatively associated with consumer intention and actual online purchases (Adjei et al., 2010). Therefore, we propose:

Hypothesis 1 (*The Live Chat Usage Hypothesis*). The live chat usage of consumers has a positive effect on their purchase decision.

3.2. Effect of seller behavior in live chat usage on the purchase decision of consumers

As suggested by literature on both information provider's behavior (Weiss et al., 2008; Song and Zinkhan, 2008) and IS usage measurement (Venkatesh and Davis, 2000), seller behavior, including time of response, number of replies, and type of information may influence subsequent behavior of consumers. We first discuss the effect of time of response and number of replies and subsequently explore the type of information.

From the perspective of product information, a prompt response from sellers helps information seekers to structure their specific problems immediately and acquire a comprehensive understanding of products (Weiss et al., 2008). Such timeliness consequently attenuates the negative effect of uncertainty. From the perspective of presence, telepresence theory posits that speed is one of the medium structures influencing the users' sense of telepresence and medium usage (Steuer, 1992). This theory predicts that users will perceive increased interactivity in a medium where information providers quickly respond to their requests (Song and Zinkhan, 2008). Interactivity can help users form trust toward sellers (Basso et al., 2001), which reduces uncertainty (Morgan and Hunt, 1994). By contrast, slow communication is a lack of responsiveness, which frustrates users (Andrews and Haworth, 2002).

As for the number of replies, frequent responses from sellers correspond to much information related to product quality and service (Weiss et al., 2008). This rich information can help mitigate the information asymmetry between consumers and sellers/products, thereby reducing product uncertainty. In addition, detailed responses from sellers can help consumers effectively understand products, enabling them to think of raising follow-up questions. The interactive process enhances consumers' sense of presence, which contributes to product uncertainty reduction (Kang et al., 2014). Reduced product uncertainty further leads to a purchase decision (Pavlou et al., 2007). Therefore, we propose:

Hypothesis 2a (*The Reply Speed Hypothesis*). In live chat usage, the timely reply from sellers has a positive effect on consumer purchase decisions.

Hypothesis 2b (*The Reply Frequency Hypothesis*). In live chat usage, the number of replies from sellers has a positive effect on consumer purchase decisions.

Literature suggests that type of information is related to product uncertainty of consumers (Adjei et al., 2010). Communication between consumers and sellers is not simply bilaterally established as message flows. This aspect requires that subsequent messages refer to previous ones, and must be relevant to the questions of consumers (Song and Zinkhan, 2008). The literature on uncertainty shows that although seller uncertainty is mitigated, consumers still face product quality and product fit uncertainties when shopping online (Dimoka et al., 2012; Hong and Pavlou, 2014). Considering these issues relevant to the concerns of consumers, this study classifies information into product quality- and product fit-related contents.

According to the definition of *product quality uncertainty* (Dimoka et al., 2012; Ghose, 2009), product quality-related contents are defined as information related to product attributes, such as product condition and quality. Apart from information in a webpage that displays products and feedback ratings, detailed information provided by sellers on product attributes according to personalized questions of consumers can help them understand product quality. This personalized information enables the accurate prediction of the future performance of a product and reduction of product quality uncertainty.

This study extends Hong and Pavlou (2014), defining product fit-related contents as information that helps consumers determine whether the attributes of a product match their preferences. Consumers believe that sellers should make suggestions or share their insights regarding personalized questions (Andrews and Haworth, 2002). Sellers who can recommend products to consumers to match their preference through live chat can help consumers determine the fit between product attributes and their own preferences. Such an action will consequently reduce product fit uncertainty.

Apart from providing relevant information, product quality- and fit-related contents can also reduce product uncertainty by increasing presence. First, consumers can sense the willingness of sellers to assist when they give attention to the concern of consumers regarding product quality and fit. Such experience provides consumers an impression of warmth and friendliness, thereby narrowing the perceived distance between consumers and sellers. Second, the relevance of these contents to personalized questions of consumers also increases their perceived interactivity. Hence, the sense of presence of consumers is increased, which will remove product uncertainty. So reduced product uncertainty raises the intention of consumers to purchase, which leads to an actual purchase decision. Therefore, we propose:

Hypothesis 2c (*The Product Quality-Related Contents Hypothesis*). Mentioning product quality-related contents in live chat has a positive effect on consumer purchase decisions.

Hypothesis 2d (*Product Fit-Related Contents Hypothesis*). Mentioning product fit-related contents in live chat has a positive effect on consumer purchase decisions.

4. Research methodology

4.1. Data collection

Data from an online skin care product store in Taobao.com were collected. The store opened in Taobao.com in 2009, and it mainly sells Clinique skin care products. We obtained data about 29,801 consumers who visited this online store via computers from June 1, 2015 to December 8, 2015, including the webpages they viewed, and the time spent on each webpage, as well as their purchase decisions and live chat usage. We also collected information about 6517 live chat conversations with 96,422 postings (questions of consumers and responses of sellers) in both computers and mobile devices during the same period. The data included the content, timestamp for each posting, and information of users who use live chat.

4.2. Model specification and key variables

The logit model is conceptually simple and frequently used in consumer behavior research in marketing and IS field (Olbrich and Holsing, 2011; De et al., 2013). Such model is typically used for situations in which one needs to predict one of only two possible values representing success and failure based on values of predictors. Olbrich and Holsing (2011) employed a logit model to analyze 2.73 million online visiting sessions of consumers and found that social shopping features exert a significant effect on consumers' decision of a click-out. De et al. (2013) checked nearly 35,000 transactions and used a logit model to estimate the influence of web technologies on the return decision of consumers. As our dependent variable is binary (purchase or not purchase), we propose logit models to demonstrate how live chat usage and seller behavior will influence the purchase decision of consumers. We formulate Model 1 to test the Live Chat Usage Hypothesis (H1) and Model 2 to test the Reply Product Hypotheses (H2a–H2d).

$$\begin{aligned} logit[P(Purchase_i = 1|X_i)] = \beta_0 + \beta_1 * LiveChatUse_i + \beta_2 * AvgWord_i \\ + \beta_3 * AvgPic_i + \beta_4 * AvgReview_i \\ + \beta_5 * AvgRating_i + \sum \beta * \text{other controls}_i + \epsilon_i \end{aligned} \quad (1)$$

$$\begin{aligned} logit[P(Purchase_j = 1|X_j)] = \beta_0 + \beta_1 * Speed_j + \beta_2 * Frequency_j \\ + \beta_3 * Fit_j + \beta_4 * Quality_j + \beta_5 * Service_j \\ + \sum \beta * Other\ Controls_j + \epsilon_j \end{aligned} \quad (2)$$

In Model 1, $Purchase_i$ is defined as whether a consumer purchased a product or not, which takes a value of 1 if consumer i successfully paid for an order, and 0 otherwise. $P(Purchase_i = 1|X_i)$ is the probability of consumer i purchasing the product for the given data. X_i is a vector of independent and control variables. For the independent variable, live chat usage ($LiveChatUse_i$) indicates whether a consumer used live chat, which takes a value of 1 if consumer i used live chat to start a conversation with a seller, and 0 otherwise. Considering that product information and user-generated contents may contribute to consumer purchase decision (Chevalier and Mayzlin, 2006; Lu et al., 2016), we control the following variables: $AvgWord_i$ is the average number of description words of products consumer i viewed. $AvgPicture_i$ is the average number of pictures of products consumer i viewed. $AvgReview_i$ is the average number of reviews of products consumer i viewed. $AvgRating_i$ is the average rating of the products consumer i viewed. Apart from the control variables we mentioned, other control variables (including product features and browsing behavior of consumers) may also affect their purchase decisions (Olbrich and Holsing, 2011; Tan et al., 2016). Thus, they are also controlled. (See Appendix A) In addition, ϵ_i is the random error.

In Model 2, $Purchase_j$ is defined as whether the consumer who participated in conversation j purchased a product. It takes a value of 1 if a consumer in conversation j successfully paid for an order, and 0 otherwise. $P(Purchase_j = 1|X_j)$ is the probability of consumer purchasing the product in conversation j for the given data. X_j is a vector of independent and control variables. For the independent variables, $Speed_j$ is the time interval between the first posting of the consumer and reply of the seller in conversation j. $Frequency_j$ is the number of replies from the seller to the questions of the consumer in conversation j. $Quality_j$ is product quality-related contents defined as whether conversation j contains information related to product attributes, such as product condition and quality. Fit_j is product fit-related contents defined as whether conversation j contains information that helps consumers determine whether the attributes of a product match their preferences. Apart from the main variables in the Reply and Product Hypotheses (H2a–H2d), we also control for consumers' information of live chat usage and online shopping as well as product features. (See Appendix B) ϵ_j is the random error.

We obtained $Quality_j$ and Fit_j in the following procedures. First, we randomly selected 437 conversations, which contained 6476 postings. Second, we manually coded these 6476 postings into one of four groups: product quality-, product fit-, service quality-related contents, and others. Service quality (e.g., product return, gifts, shipping policy) may influence the purchase decision of consumers (Valvi and Fragkos, 2012). Thus, we also included service quality in the classification. Third, we trained the algorithms of *support vector machine* (SVM) and *naive Bayes* (NB) using 89.6% (i.e., 5800) of the coded postings as input to build the two classification models. Fourth, we used the two models to classify the remaining 10.4% (i.e., 676) of the postings. The accuracy level of SVM is 82.98%, which is higher than that of NB (78.01%). Finally, SVM was applied to classify the remaining 89,946 postings into four groups. $Quality_j$ takes a value of 1 as long as one posting of the seller in conversation j is classified into the product quality-related contents group; and otherwise, a value of 0 is taken. Fit_j and $Service_j$ refer to product fit and service quality-related contents, respectively, and their measurements are similar to $Quality_j$.

5. Data analysis and results

5.1. Descriptive statistics

Table 1 shows that only 3.0% of 29,801 consumers purchased products, so the purchase conversion rate online is low. In total, 2.7% of consumers chatted with sellers. The means of average words and pictures for each product consumers viewed are 1,220.8 and 7.6, respectively. The means of average reviews and ratings for each product that consumers viewed are 24.2 and 4.8, which suggests a fairly high evaluation. Table 2 shows the correlations among the main variables in Model 1.

The descriptive statistics of 6517 live chat conversations in Table 3 show that 2381 purchase records existed after live chat usage. The seller replied 7.6 times and provided the first reply in 677.5 s in each conversation on average. The longest time for $speed_j$ was 80,911 s before the seller replied to the first question of consumers. The seller provided product fit-related contents in 30.7% conversations, whereas product quality-related contents were provided in 51.7%. Table 4 shows the correlations among variables that measure the behavior of the seller.

5.2. Hypothesis testing

5.2.1. Effect of live chat usage on consumer purchase decision

Empirical results of Model 1. Table 5 reports the estimated results for Model 1 (the full table, including other control variables, is in Appendix C). This model significantly fits better than a null model with only an intercept and no covariates (χ^2 (12) = 2,013.53, $p < .01$)). The pseudo-R^2 is 27%. The commonly-used pseudo-R^2 statistic is employed to measure fit of logit model, that is, the McFadden's R^2. This statistic refers to the percentage comparison between the log likelihood of the fitted model and the null model. Measures of pseudo-R^2 do not explain the variance in the same way as the R^2 coefficient in linear regression. In addition, measures of pseudo-R^2 are not as high as R^2 in linear regression, in which a pseudo-R^2 between 0.2 and 0.4 indicates a good model fit according to McFadden (1979).

Table 5 shows that live chat, which occurred between consumers and sellers, has a positive coefficient ($\beta = 3.431$, $p < .01$). Moreover, the marginal effect of live chat on purchase decision is 0.30 (using the mfx command in Stata), that is, live chat usage is associated with a 0.30 increase in the probability of the purchase

Table 1
Descriptive statistics for all consumers.

Variable	Mean	Std. Dev.	Min	Max
$Purchase_i$	0.03	0.17	0	1
$LiveChatUse_i$	0.03	0.16	0	1
$AvgWord_i$	1220.83	744.38	0	3711
$AvgPic_i$	7.60	4.25	0	18
$AvgReview_i$	24.21	24.35	0	130
$AvgRating_i$	4.79	0.13	4.2	5

Table 3
Descriptive statistics of variables related to live chat conversations.

Variable	Mean	Std. Dev.	Min	Max
$Purchase_j$	0.36	0.48	0	1
$Frequency_j$	7.63	9.72	0	122
$Speed_j$	677.54	3,740.75	0	80,911
Fit_j	0.31	0.46	0	1
$Quality_j$	0.52	0.50	0	1
Service	0.59	0.49	0	1

Table 2
Correlation among the main variables in Model 1.

Variables	$Purchase_i$	$LiveChatUse_i$	$AvgWord_i$	$AvgPic_i$	$AvgReview_i$	$AvgRating_i$
$Purchase_i$	1					
$LiveChatUse_i$	0.413	1				
$AvgWord_i$	0.058	0.043	1			
$AvgPic_i$	0.002	0.020	0.319	1		
$AvgReview_i$	0.022	0.020	0.311	0.123	1	
$AvgRating_i$	0.015	0.026	0.124	0.256	−0.293	1

Table 4
Correlation among the main variables in Model 2.

Variable	$Purchase_j$	$Frequency_j$	$Speed_j$	Fit_j	$Quality_j$	$Service_j$
$Purchase_j$	1					
$Frequency_j$	0.183	1				
$Speed_j$	−0.150	−0.157	1			
Fit_j	0.160	0.516	−0.129	1		
$Quality_j$	0.085	0.433	−0.188	0.452	1	
$Service_j$	0.348	0.231	−0.175	0.029	−0.046	1

Table 5
Estimation results for the effect of live chat use on consumer purchase decision.

Variables	Coefficient	SE.
$LiveChatUse_i$	3.431***	0.093
Control Variables		
$AvgWord_i$	3.16E−04***	5.98E−05
$AvgPic_i$	0.033**	0.015
$AvgReview_i$	0.002	0.002
$AvgRating_i$	−0.224	0.381
Other Controls		
Intercept	−3.111*	1.820
N	29,801	
Log-likelihood	−2,928.65	
χ^2 (12)	2,013.53***	
Pseudo R^2	0.271	

* $p < .1$
** $p < .05$.
*** $p < .01$.

decision of consumers compared with the non-existence of live chat usage. Therefore, the Live Chat Usage Hypothesis (H1) is supported.

Product descriptions (descriptive words and pictures) and user-generated contents (ratings and number of product reviews) are also examined (Dimoka et al., 2012; Chevalier and Mayzlin, 2006; Lu et al., 2016). The effects of average words and pictures of the products are significant and positive (β = 3.16E−04, $p < .01$; β = 0.033, $p < .05$, respectively), and the marginal effects of average words and pictures are 5.00E−06 and 5.27E−04, respectively. The effects of average rating and number of reviews are insignificant. The descriptive statistics show that the average rating of products is relatively high, with a mean of 4.789 and small standard deviation of 0.132. Most reviews and ratings are positive and similar to each other. These reviews and ratings may not be helpful to consumers because consumers cannot rely on them to distinguish one product from another. Therefore, consumers do not depend on reviews and ratings when making a purchase decision.

As for the other control variables, consumers who spend more time on product pages, view more product webpages, look around for men's products or come from a wealthier city are more eager to conduct purchase decision than those who do not. Consumers who view new arrival products and products with sub category are less likely to purchase. Product price influences the purchase decision of consumers negatively.

However, $LiveChatUse_i$ is an endogenous variable potentially related to some unobserved variables. In the following section, we attempt to alleviate the endogenous issue by measuring other related variables, and use the *coarsened exact matching* (CEM) *method* to improve the estimation of causal effects.

Robustness check. One of the most important unobserved variables in our context is consumers' desire to purchase, that is, consumers with a strong desire to buy tend to use live chat. Accordingly, we used the following variables to measure consumers' urgency to purchase.

$DiversityDegree_i$ is the degree to which consumer i is distracted from his or her initial shopping goal. The more diverse the products a consumer view, the weaker the initial purchasing desire she has, and this consumer unlikely makes a purchase. This variable is measured by the number of different categories of viewed products. $SimilarityDegree_i$ is the degree to which consumer i concentrates on his or her desire to buy. We assume that the more similar products a consumer browse, the more likely he or she will finally make a purchase decision. $SimilarityDegree_i$ is measured by the number of browsed products belonging to the category of the first product he or she viewed in the store that day. $DirectBrowse_i$ is used to determine how consumer i obtains access to the online store. This variable takes a value of 1 if consumer i is directed by links in shopping favorite lists, shopping collections, or purchase history to access the store. These consumers have once visited the online store, and they have additional knowledge regarding the store and its products (Meyer, 1982). $DirectBrowse_i$ takes a value of 0 if consumer i accessed the store by searching related products or clicking banner advertisement links. These consumers may have no shopping experience in this store, and are unfamiliar about the quality of products sold, suggesting that they are in a high level of uncertainty. Table 6 reports the estimated results for Model 1 with three new variables.

The estimated results in Table 6 show that less diversity of viewed products, more viewed products belonging to the category of first viewed product, and having once visited the online store are all associated with increased purchase probability (the full table, including other control variables, is in Appendix D) After controlling for consumers' urgency to purchase, the effect of live chat use on purchase decision remains positive ($\beta = 3.418$, $p < .01$). The effect of the average number of pictures is insignificant. The effects of the rest control variables remain stable.

Thereafter, we used CEM to compare the purchase decisions of those with and without live chat (treated and control consumers, respectively). The CEM method uses an exact match on coarsened variables and runs the analysis on the uncoarsened, matched data.

Table 6
Effect of live chat usage controlling for consumers' urgency to purchase.

Variables	Coef.	SE
$LiveChatUse_i$	3.418***	0.096
Control Variables		
$AvgWord_i$	2.92E-04***	6.58E-05
$AvgPic_i$	0.013	0.016
$AvgReview_i$	0.002	0.002158
$AvgRating_i$	−0.413	0.419
$DiversityDegree_i$	−0.096**	0.040
$SimilarityDegree_i$	0.052*	0.031
$DirectBrowse_i$	1.182***	0.085
Other Controls		
Intercept	−2.453	2.078
N	29,801	
Log-likelihood	−2,819.10	
χ^2 (15)	1,978.34***	
Pseudo R^2	0.2985	

* $p < .1$.
** $p < .05$.
*** $p < .01$.

This method does not require the researcher to set the size of the matching solution ex ante, but check for balance *ex post*, which is a requirement in most common matching methods (Blackwell et al., 2009). The basic idea is that the different purchase decisions between the control and treated consumers after matching are due to whether they are treated (i.e., live chat usage).

Factors influencing the choice of consumers to use live chat must be determined before performing CEM (Subramanian and Overby, 2017). For the 797 consumers who used live chat, 65.12% began using it after they finished browsing or when they were browsing the last webpage. Thus, their browsing experience in the online store would have influenced the live chat usage of consumers. In this study, all product/webpage feature variables, urgency to purchase, and other control variables were used as determinants of live chat. We then matched treated and control consumers with all these variables.

Continuous control variables are coarsened into bins of each 10th percentile width. For dummy variables, we utilized exact matching. To ensure that the procedure resulted in comparable matches, we examined the balance between the treated and control groups. We first calculated the means of control variables for the treated and control consumers in each stratum in the matched sample. We then used a *t*-test to examine whether the means of these strata differ significantly between the two groups. Table 7 shows the results. (The full table, including other control variables, is in Appendix E).

In Table 7, $AvgWord_i$ and $SimilarityDegree_i$ show significant differences after matching, but the difference contains minimal practical significance (difference in means: $AvgWord_i$, 0.14; $SimilarityDegree_i$, 0.06). Overall, the matching results show that treated and control consumers are quite comparable after matching.

Using the matched sample, we estimated a logit model to test the effect of live chat usage:

$$logit[P(Purchase_i = 1|X_i)] = \beta_0 + \beta_1 LiveChatUse_i + \epsilon_i \quad (3)$$

We fit the above model with the matched sample using weighted regression, with the weights provided by the CEM procedures (Iacus et al., 2012). Table 8 shows the results, which are similar to the previous results in Tables 5 and 6.

Table 8
Effect of live chat use with matched sample.

Variables	Coef
$LiveChatUse_i$	2.80 (0.41)***
Intercept	−3.89 (0.38) ***
N	5144
Log likelihood; χ^2 (1)	−606.49; 47.62***

*** $p < .01$.

5.2.2. Effect of seller behavior in live chat usage on consumer purchase decision

Empirical results of Model 2. To address the endogenous problem that the seller may sense consumers' strong purchase desire and thus behave differently, we also controlled for consumers' urgency to purchase. $ChatDiversityDegree_j$ is measured by the number of product categories mentioned in conversation *j*. $ChatSimilarityDegree_j$ is the degree to which the consumer concentrates on their desire to buy, and is measured by the number of discussed products belonging to the category of the first product he or she mentioned in conversation *j*. $RptDialogue_j$ refers to whether conversation *j* happens between sellers and a consumer who has previously communicated with sellers within a week. $RptDialogue_j$ takes a value of 1 if the consumer once had a conversation with the seller, and 0 if otherwise.

Table 9 reports the estimation results for Model 2 (the full table, including other control variables, is in Appendix F). The time of the first reply is negatively related to the purchase decision of consumers ($\beta = -3.50E-05$, $p < .01$). Thus, speed of the first reply positively affects the purchase decision of consumers.

Table 9
Estimation results for the effect of seller behavior in live chat usage.

Variable	Coef	SE
$Frequency_j$	0.013***	0.004
$Speed_j$	−3.50E−05***	3.86E−06
Fit_j	0.662***	0.076
$Quality_j$	0.136*	0.073
Control Variables		
$Service_j$	1.506***	0.072
$ChatDiversityDegree_j$	−0.150*	0.084
$ChatSimilarityDegree_j$	−0.046	0.029
$RptDialogue_j$	0.569***	0.105
Other Controls		
Intercept	−3.451*	1.803
N	6517	
Log-likelihood	−3,468.040	
χ^2 (19)	1,187.18***	
Pseudo R^2	0.189	

* $p < .1$.
*** $p < .01$.

Table 7
Means comparison for treated and control consumers for Hypothesis 1.

Variable	# Strata	# Treated Consumers	# Control Consumers	Mean: Treated Consumers	Mean: Control Consumers	Difference in Means	*t*-stat
Panel A. Before matching (full sample)							
$AvgPic_i$	–	797	29,004	8.12	7.59	0.53	3.48
$AvgWord_i$	–	797	29,004	1,414.51	1,215.5	199.01	7.45
$AvgReview_i$	–	797	29,004	27.21	24.13	3.08	3.52
$AvgRating_i$	–	797	29,004	4.81	4.78	0.21	4.53
$DiversityDegree_i$	–	797	29,004	1.55	1.74	−0.19	−2.64
$SimilarityDegree_i$	–	797	29,004	3.89	1.51	2.38	34.7
$DirectBrowse_i$	–	797	29,004	0.36	0.23	0.13	8.76
Panel B. After matching (matched sample)							
$AvgPic_i$	168	223	4921	8.76	8.77	−0.001	−1.99
$AvgWord_i$	168	223	4921	1,428.31	1,428.18	0.14	2.07
$AvgReview_i$	168	223	4921	26.65	26.64	0.01	1.88
$AvgRating_i$	168	223	4921	4.81	4.8	0.001	1.38
$DiversityDegree_i$	168	223	4921	1.05	1.05	0	0
$SimilarityDegree_i$	168	223	4921	1.44	1.38	0.06	3.23
$DirectBrowse_i$	168	223	4921	0.16	0.16	0	0

Therefore, the Reply Speed Hypothesis (H2a) is supported. Frequency of seller replies positively influences the purchase decision of consumers ($\beta = 0.013$, $p < .01$). Therefore, the Reply Frequency Hypothesis (H2b) is supported. Product quality- and fit-related contents significantly affect purchase decision regarding information type, supporting both the Product Quality-Related Contents Hypothesis (H2c) and the Product Fit-Related Contents Hypothesis (H2d).

Next, consider the results shown in Table 9.

Product fit-related contents greatly affected the decisions of consumers compared with product quality-related contents. It makes sense because consumers can obtain product quality information from other sources such as website description, whereas consumers can only obtain adequate information on fit-related question from chatting with the seller.

Product descriptions provided on webpages are too general to satisfy the personal issues of consumers. Reviews and comments regarding a product from other individuals can occasionally be helpful for the fit question of consumers, only if the following conditions hold. First, former consumers had the same product-fit related uncertainty, and with similar skin condition (in our case) as this consumer. Second, these former consumers were willing to provide feedback after trying the products. Third, this consumer is willing to spend time looking for a specific answer from thousands of reviews. Chatting with the seller is more convenient and efficient than searching specific product fit-related contents in the reviews because sellers have accumulated fit information from previous cases of consumers. Then, they act as agents transmitting the fit information of former consumers (which may not be written in reviews) to the current consumer. Hence, the consumer can acquire information to reduce product fit-related uncertainty immediately.

Among the other control variables, consumers with high membership rank (i.e., more expenditure in the e-commerce site than others), who mention products with lower price or more pictures, and who previously chatted with sellers within a week or male consumers are likely to make a purchase decision.

Robustness check. Other unobserved variables may still exist apart from the urgency of consumers to purchase. Thus, we used the CEM method to compare the purchase decision of those who want to obtain product fit-related contents from sellers (treated consumers) with those who do not obtain product fit-related contents (control consumers). Refer to Appendix G for details. The results of the comparison between consumers who obtain product quality-related contents from sellers with those who do not discuss product quality-related contents are quite similar. Thus, the results are not presented.

We fit Model 2 with the matched sample using weighted regression. Table 10 shows the results. (The full table, including other control variables, is in Appendix G) The results are similar to the previous results in Table 9. Purchase decisions were positively affected by product fit-related contents mentioned in the chat ($\beta = 0.591$, $p < .01$). Product quality-related contents also significantly influenced purchase decisions ($\beta = 0.157$, $p < .1$). The effects of other variables are quite similar too.

Table 10
Effect of seller behavior in live chat usage with matched samples.

Variable	Coef	SE
$Frequency_j$	0.010**	0.005
$Speed_j$	−3.40E−05***	4.57E−06
Fit_j	0.591***	0.084
$Quality_j$	0.157*	0.082
Control Variables		
$Service_j$	1.701***	0.078
$ChatDiversityDegree_j$	−0.077	0.095
$ChatSimilarityDegree_j$	−0.060**	0.030
Other Controls		
Intercept	−2.972	2.000
n	6014	
Log-likelihood	−3,213.491	
χ^2 (14)	937.90***	
Pseudo R^2	0.187	

* $p < .1$.
** $p < .05$.
*** $p < .01$.

6. Conclusion

6.1. Key findings

This study analyzed the relationship between live chat usage and consumer purchase decisions using data from a skin care product store. Our results confirm that consumers who chat with sellers tend to make a purchase. During live chat, the behavior of sellers, including their fast and frequent replies, as well as the product quality- and fit-related contents provision, increase the possibility that consumers will make purchase decisions.

This study makes the following contributions. First, our research addressed and confirmed the importance of live chat on the purchase decisions of consumers. Apart from webpage features and product forums (Hong and Pavlou, 2014; Dimoka et al., 2012), our study explains that live chat acts as an important medium in uncertainty reduction and relationship establishment by providing information and increasing presence. Compared with other relevant literature on live chat (Jiang et al., 2010; Lu et al., 2016), our study goes beyond the dependent variable of purchase intention. Instead, consumer purchase decisions were used to reflect the actual decision of consumers after live chat. This research aids in understanding the antecedents of online shopping decision with additional comprehensive insights.

Second, our research has explored live chat usage through sellers to influence consumer purchase decisions, extending the studies on the effect of live chat (Tan et al., 2016; Ou et al., 2014). Prior literature focused on the effect of live chat (Ou et al., 2014; Jiang et al., 2010; Kang et al., 2014), and neglecting approaches regarding live chat usage may make a difference. As suggested by the literature on IS usage (Burton-Jones and Straub, 2006), and the behavior of information providers, our study develops measures for live chat usage. We also conduct intensive study to prove that of the measures for live chat usage, response speed, response frequency and information type are key factors that determine consumers' final decision.

Finally, our research extended and provided evidence for uncertainty reduction theory. We analyze the information type in live chat based on the classification of uncertainty in marketing literature into product quality- and product fit-related contents (Hong and Pavlou, 2014). We disclose that although both product quality- and product fit-related contents can positively affect consumer purchase decision, product fit-related contents are significant, underlining the urgency to solve personalized questions of consumers, and matching their preference in the online marketplace.

6.2. Managerial implications

This study establishes the following managerial implications for sellers to use live chat. First, live chat between consumers and sellers is important, especially for consumers who have never been to the store before. Sellers can take advantage of monitoring tools (e.g., Shengyicanmou provided by Taobao.com) to identify new consumers and allow an experienced staff to communicate with them. Second, sellers should reply to consumers as quickly as possible because a rapid reply can promptly provide information before the uncertainty results in a negative effect. Moreover, a quick response can leave consumers with a good impression of high-quality service, thereby leading to a pleasant online shopping experience.

Third, sellers can convey rich and comprehensive information to consumers by replying frequently, thereby helping reduce uncertainty and transform information seekers into consumers. Finally, sellers can provide more contents regarding the fit between products and the preference of a consumer, as well as product quality when the consumer starts several conversations. A conversation after another in several days indicates that the consumer still seriously considers purchasing the product. In this situation, additional product fit-related contents may look trustworthy, which tend to convince a consumer about the suitability of the product.

6.3. Limitations

This study also has some limitations. First, though our models were empirically tested to be valid, other explanatory variables that we did not control for may possibly have exerted influence on the purchase decisions of consumers. These include such things as their demographic information (income, age, and personalities), context factors (website technology, the existence of competing products and brands), and social factors interaction of consumers within the brand community) (Casper 2007; De et al., 2013; Naylor et al., 2012; Mallapragada et al., 2016; Adjei et al., 2010; Moslehpour et al., 2016). However, we did not obtain these variables because of resource and data limitation. We suggest researchers to consider these variables in future research.

Second, the data analyzed in this study came from a store selling skin care products, which are typical experience goods (Nelson, 1974). Consumers may rely on different kinds of information when they buy other products. Thus, users of our findings should be cautious in applying them in other contexts.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (grant number 71490721 and 71272028).

Appendix A. Control variables for Model 1

Apart from the variables stated in Section 4.2, product features and browsing behavior of consumers may also affect consumer purchase decision (Olbrich and Holsing, 2011; Tan et al., 2016). Thus, these factors are controlled in this study. $AvgPrice_i$ refers to the average number of prices in product pages that consumer i viewed. $SubCategory_i$ takes a value of 1 if any of the viewed products have different specifications for consumer i to choose, such as different colors and sizes of the blusher and moisturizing lotion, respectively. Otherwise, $SubCategory_i$ takes a value of 0. $NumNew_i$ refers to the number of viewed new arrival products. $AvgTime_i$ denotes the average time consumer i spent in each of the browsed products. $NumPrdctPage_i$ refers to the number of products viewed by consumer i. $Gender_i$ refers to the suggested gender of products consumer i viewed, which takes a value of 1 if products consumer i viewed are for women and 0 if otherwise. $GDPRank_i$ denotes the GDP rank of city, where consumer i conducted online browsing behavior in 2015 according to the China National Bureau of Statistics (Yue, 2015).

Appendix B. Control variables for Model 2

Apart from the main variables in Hypotheses 2a-2d, live chat usage and online shopping information consumers obtained are controlled, as well as product features. $CreditLevel_j$ refers to the credit of the consumer when conversation j occurred, which was evaluated by former sellers. $MemberLevel_j$ refers to the membership rank of the consumer when conversation j happened, which was associated with their total expenditure in Taobao.com. $ActiveLevel_j$ denotes the activeness of the consumer in live chat tool usage when conversation j took place, which was associated with the time spent in live chat dialogues from signing up. $WangGender_j$ refers to the gender of the consumer. $ChatPrdctWord_j$ refers to the average number of words in the webpages of products which were mentioned in conversation j. $ChatPrdctPic_j$ is the average number of pictures in the webpages of products which were mentioned in conversation j. $ChatReview_j$ refers to the average number of reviews of products which were mentioned in conversation j. $ChatRating_j$ denotes the average rating of products which were mentioned in conversation j. $ChatPrice_j$ refers to the average price of products which were mentioned in conversation j. $ChatNew_j$ refers to the number of mentioned products in conversation j, which were also new arrival products. $PrdctMentioned_j$ was a binary variable, taking a value of 1 if a consumer mentioned a product in conversation j.

Appendix C

Table C1.

Table C1
Full Estimation Results for Model 1.

Variables	Coefficient	SE.
$LiveChatUse_i$	3.431***	0.093
Control Variables		
$AvgWord_i$	3.16E-04***	5.98E-05
$AvgPic_i$	0.033**	0.015
$AvgReview_i$	0.002	0.002
$AvgRating_i$	−0.224	0.381
$AvgPrice_i$	−0.001*	3.90E-04
$AvgTime_i$	7.07E-05***	1.73E-05
$SubCategory_i$	−0.662***	0.135
$NumNew_i$	−0.176***	0.048
$NumPrdctPage_i$	0.150***	0.012
$Gender_i$	−0.537***	0.112
$GDPRank_i$	−0.005***	0.001
Intercept	−3.111*	1.820
N	29,801	
Log-likelihood	−2,928.65	
χ^2 (12)	2,013.53***	
Pseudo R^2	0.2713	

* $p < .1$.
** $p < .05$.
*** $p < .01$.

Appendix D

Table D1

Table D1
Effect of Live Chat Usage While Controlling for Consumer Urgency to Purchase.

Variables	Coef	SE
$LiveChatUse_i$	3.418***	0.096
Control Variables		
$AvgWord_i$	2.92E−04***	6.58E−05
$AvgPic_i$	0.013	0.016
$AvgReview_i$	0.002	0.002158
$AvgRating_i$	−0.413	0.419
$AvgPrice_i$	−3.99E−04	3.70E−04
$AvgTime_i$	5.96E−05***	1.81E−05
$SubCategory_i$	−0.502***	0.137
$NumNew_i$	−0.133***	0.048
$NumPrdctPage_i$	0.115***	0.024
$Gender_i$	−0.442***	0.116
$GDPRank_i$	−0.006***	0.001
$DiversityDegree_i$	−0.096**	0.040
$SimilarityDegree_i$	0.052*	0.031
$DirectBrowse_i$	1.182***	0.085
Intercept	−2.453	2.078
N	29,801	
Log-likelihood	−2,819.10	
χ^2 (15)	1,978.34***	
Pseudo R^2	0.2985	

* $p < .1$.
** $p < .05$.
*** $p < .01$.

Appendix E

Table E1

Table E1
Comparison of Means for Treated and Control Consumers for Hypothesis 1.

Variable	# Strata	# Treated Consumers	# Control Consumers	Mean Treated Consumers	Mean Control Consumers	Difference in Means	*t*-Stat
Panel A. Before matching (full sample)							
$AvgTime_i$	–	797	29,004	403.77	183.12	220.65	5.48
$AvgPic_i$	–	797	29,004	8.12	7.59	0.53	3.48
$AvgWord_i$	–	797	29,004	1414.51	1215.5	199.01	7.45
$AvgReview_i$	–	797	29,004	27.21	24.13	3.08	3.52
$AvgPrice_i$	–	797	29,004	133.2	129.56	3.64	0.84
$AvgRating_i$	–	797	29,004	4.81	4.78	0.21	4.53
$SubCategory_i$	–	797	29,004	0.25	0.32	0.07	4.05
$NumNew_i$	–	797	29,004	0.79	0.31	0.47	18.44
$NumPrdctPage_i$	–	797	29,004	5.26	1.78	3.48	34.81
$Gender_i$	–	797	29,004	0.79	0.87	−0.08	−7.02
$GDPRank_i$	–	797	29,004	32.18	36.45	−4.27	−3.17
$DiversityDegree_i$	–	797	29,004	1.55	1.74	−0.19	−2.64
$SimilarityDegree_i$	–	797	29,004	3.89	1.51	2.38	34.7
$DirectBrowse_i$	–	797	29,004	0.36	0.23	0.13	8.76
Panel B. After matching (matched sample)							
$AvgTime_i$	168	223	4921	491.17	394.91	96.25	0.93
$AvgPic_i$	168	223	4921	8.76	8.77	−0.001	−1.99
$AvgWord_i$	168	223	4921	1,428.31	1,428.18	0.14	2.07
$AvgReview_i$	168	223	4921	26.65	26.64	0.01	1.88
$AvgPrice_i$	168	223	4921	176.25	176.06	0.19	1.24
$AvgRating_i$	168	223	4921	4.81	4.8	0.001	1.38
$SubCategory_i$	168	223	4921	0.41	0.41	0	0
$NumNew_i$	168	223	4921	0.4	0.38	0.02	1.8
$NumPrdctPage_i$	168	223	4921	1.45	1.39	0.06	3.23
$Gender_i$	168	223	4921	0.81	0.81	0	0
$GDPRank_i$	168	223	4921	32.28	33	−0.71	−1.6
$DiversityDegree_i$	168	223	4921	1.05	1.05	0	0
$SimilarityDegree_i$	168	223	4921	1.44	1.38	0.06	3.23
$DirectBrowse_i$	168	223	4921	0.16	0.16	0	0

Appendix F

Table F1

Table F1
Estimation results for the effect of communication quality and contents.

Variable	Coef	SE
$Frequency_j$	0.013***	0.004
$Speed_j$	−3.50E-05***	3.86E-06
Fit_j	0.662***	0.076
$Quality_j$	0.136*	0.073
Control Variables		
$Service_j$	1.506***	0.072
$ActiveLevel_j$	2.92E-06*	1.61E-06
$MemberLevel_j$	0.228***	0.039
$CreditLevel_j$	0.023	0.021
$WangGender_j$	−0.311***	0.088
$ChatPrdctWord_j$	9.00E-06	5.76E-05
$ChatPrdctPic_j$	0.234***	0.013
$ChatReview_j$	0.003	0.002
$ChatRating_j$	−0.115	0.367
$ChatPrice_j$	−0.001***	3.02E-04
$ChatNew_j$	−0.131*	0.072
$PrdctMentioned_j$	−0.022	0.078
$ChatDiversityDegree_j$	−0.150*	0.084
$ChatSimilarityDegree_j$	−0.046	0.029
$RptDialogue_j$	0.569***	0.105
Intercept	−3.451*	1.803
N	6517	
Log-likelihood	−3,468.040	
χ^2 (19)	1,187.180***	
Pseudo R^2	0.189	

* $p < .1$
*** $p < .01$

Appendix G. Effect of communication quality and contents with matched samples

To match consumers discussing product fit-related contents with sellers (treated consumers) to those who do not (control consumers), we used consumers' information on live chat usage and online shopping ($ActiveLevel_j$, $MemberLeve_j$, $CreditLevel_j$), gender ($WangGender_j$), and whether this chat is a repeat dialogue in a week ($RptDialogue_j$) as the antecedent variables of talking about product fit-related contents. We then matched treated and control consumers using all these variables.

We coarsened continuous variables ($ActiveLevel_j$ and $CreditLevel_j$) according to their distribution. They are coarsened into bins with a 10th percentile width for each. For categorical variables ($MemberLevel_j$, $WangGender_j$, $RptDialogue_j$), we used exact matching. The balance between treated and control groups were examined, and no significant difference was found. Overall, the matching results showed that treated and control consumers are quite comparable after matching. Table G1 shows the results of Model 2 with matched sample.

Table G1
Effect of Communication Quality and Contents with Matched Samples.

Variable	Coefficient	SE
$Frequency_j$	0.010**	0.005
$Speed_j$	−3.40E−05***	4.57E−06
Fit_j	0.591***	0.084
$Quality_j$	0.157*	0.082
Control Variables		
$Service_j$	1.701***	0.078
$ChatPrdctWord_j$	−3.80E−05	6.53E−05
$ChatPrdctPic_j$	0.251***	0.015
$ChatReview_j$	0.001	0.002
$ChatRating_j$	−0.13	0.412
$ChatPrice_j$	−0.001***	3.35E-04
$ChatNew_j$	−0.183**	0.077
$PrdctMentioned_j$	0.071	0.081
$ChatDiversityDegree_j$	−0.077	0.095
$ChatSimilarityDegree_j$	−0.060**	0.030
Intercept	−2.972	2.000
N	6014	
Log-likelihood	−3213.49	
χ^2 (14)	937.90***	
Pseudo R^2	0.187	

* $p < .1$.
** $p < .05$.
*** $p < .01$.

References

Adjei, M.T., Noble, S.M., Noble, C.H., 2010. The influence of C2C communications in online brand communities on customer purchase behavior. J. Acad. Mark. Sci. 38 (5), 634–653.
Alavi, M., Henderson, J.C., 1981. An evolutionary strategy for implementing a decision support system. Manage. Sci. 27 (11), 1309–1323.
Andrews, D.C., Haworth, K.N., 2002. Online customer service chat: usability and sociability issues. J. Internet Market. 2 (1), 1–20.
Basso, A., Goldberg, D., Greenspan, S., Weimer, D., 2001. First impressions: emotional and cognitive factors underlying judgments of trust e-commerce. In: Proceedings of the 3rd ACM Conference on Electronic Commerce, Tampa, Florida. ACM Press, New York, pp. 137–143.
Benbasat, I., Gefen, D., Pavlou, P.A., 2008. Special issue: trust in online environments. J. Manage. Inf. Syst. 24 (4), 5–11.
Blackwell, M., Iacus, S., King, G., Porro, G., 2009. CEM: coarsened exact matching in Stata. Stata J. 9 (4), 524–546.
Burton-Jones, A., Straub Jr, D.W., 2006. Reconceptualizing system usage: an approach and empirical test. Inf. Syst. Res. 17 (3), 228–246.
Casper, J., 2007. Sport commitment, participation frequency and purchase intention segmentation based on age, gender, income and skill level with U.S. tennis participants. Eur. Sport Manage. Q. 7 (3), 269–282.
Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. J. Market. Res. 43 (3), 345–354.
De, P., Hu, Y., Rahman, S.M., 2013. Product-oriented web technologies and product returns: an exploratory study. Inf. Syst. Res. 24 (4), 998–1010.
DeLone, W.H., McLean, E.R., 1992. Information systems success: the quest for the dependent variable. Inf. Syst. Res. 3 (1), 60–95.
Dimoka, A., Hong, Y., Pavlou, P.A., 2012. On product uncertainty in online markets: theory and evidence. MIS Q. 36 (2), 395–426.
Ghose, A., 2009. Internet exchanges for used goods: an empirical analysis of trade patterns and adverse selection. MIS Q. 33 (2), 263–291.
Hong, Y., Pavlou, P.A., 2014. Product fit uncertainty in online markets: nature, effects, and antecedents. Inf. Syst. Res. 25 (2), 328–344.
Iacus, S.M., King, G., Porro, G., Katz, J.N., 2012. Causal inference without balance checking: coarsened exact matching. Polit. Anal. 20, 1–20.
Ji, R., Meeker, M., 2005. Creating Consumer Value in Digital China. Industry Overview. Morgan Stanley, New York.
Jiang, Z., Chan, J., Tan, B.C., Chua, W.S., 2010. Effects of interactivity on website involvement and purchase intention. J. Assoc. Inf. Syst. 11 (1), 34–59.
Kang, L., Wang, X., Tan, C.H., Zhao, J.L., 2014. Understanding the antecedents and consequences of Live-Chat use in e-commerce context. In: International Conference on HCI in Business. Heraklion, Crete, Greece, pp. 504–515.
Lowry, P.B., Romano, N.C., Jenkins, J.L., Guthrie, R.W., 2009. The CMC interactivity model: how interactivity enhances communication quality and process satisfaction in lean-media groups. J. Manage. Inf. Syst. 26 (1), 155–196.
Lu, B., Fan, W., Zhou, M., 2016. Social presence, trust, and social commerce purchase intention: an empirical research. Comput. Human Behav. 56, 225–237.
Mallapragada, G., Chandukala, S.R., Liu, Qing, 2016. Exploring the effects of "what" (product) and "where" (website) characteristics on online shopping behavior. J. Market. 80 (2), 21–38.
McFadden, D., 1979. Quantitative methods for analyzing travel behaviour on individuals: some recent developments. In: Hensher, D.A., Stopher, P.R. (Eds.), Behavioral Travel Modelling. Croom Helm, London, pp. 279–318.
Meyer, R.J., 1982. A descriptive model of consumer information search behavior. Market. Sci. 1 (1), 93–121.
Ministry of Commerce of the People's Republic of China, 2017. Regular Press Conference of the Ministry of Commerce on 9th February 2017. Available at: <www.mofcom.gov.cn/article/ae/slfw/201702/20170202512614.shtml>.
Morgan, R.M., Hunt, S.D., 1994. The commitment-trust theory of relationship marketing. J. Market. 58 (3), 20–38.
Moslehpour, M., Fedosova, J., Teguh, N.M., 2016. The impact of personality traits and perception of technology towards online purchase intention of Taiwanese college students. The Conference on Theories and Practices in International Business and Conference on Management Master Theses in Southern Taiwan, Tainan, Taiwan.
Musil, S., 2009. Sold! eBay jettisons Skype in $2 billion deal. Available at: <www.cnet.com/news/sold-ebay-jettisons-skype-in-2-billion-deal/>.
Naylor, R.W., Lamberton, C.P., West, P.M., 2012. Beyond the "Like" button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. J. Market. 76 (6), 105–120.
Nelson, P., 1974. Advertising as information. J. Polit. Econ. 82 (4), 729–754.
Olbrich, R., Holsing, C., 2011. Modeling consumer purchasing behavior in social shopping communities with clickstream data. Int. J. Electr. Commerce 16 (2), 15–40.
Ou, C.X., Pavlou, P.A., Davison, R., 2014. Swift guanxi in online marketplaces: the role of computer-mediated communication technologies. MIS Q. 38 (1), 209–230.
Pavlou, P.A., Liang, H., Xue, Y., 2007. Understanding and mitigating uncertainty in online environments: a principal-agent perspective. MIS Q. 31 (1), 105–136.
Peck, J., Childers, T.L., 2003. To have and to hold: the influence of haptic information on product judgments. J. Market. 67 (2), 35–48.
Saga, V.Y., Zmud, R.W., 1994. The nature and determinants of IT acceptance, routinization and infusion. In: Levine, L., (Ed.), Diffusion, Transfer and Implementation of Information Technology, first ed., IFIP Transactions 45. North Holland, Amsterdam, Netherlands.
Short, J., Williams, E., Christie, B., 1976. The Social Psychology of Telecommunications. John Wiley and Sons, New York.
Song, J.H., Zinkhan, G.M., 2008. Determinants of perceived web site interactivity. J. Market. 72 (2), 99–113.
Steuer, J., 1992. Defining virtual reality: Dimensions determining telepresence. J. Commun. 42 (4), 73–93.
Subramanian, H., Overby, E., 2017. Electronic commerce, spatial arbitrage, and market efficiency. Inf. Syst. Res. 28 (1), 97–116.
Tan, X., Wang, Y., Tan, Y., 2016. The value of live chat on online purchase. In: Proceedings of the 37th International Conference on Information Systems, Dublin, Ireland.
Tan, X., Wang, Y., Tan, Y., 2016. The value of live chat on online purchase. Proceedings of the 37th International Conference on Information Systems, Dublin, Ireland.
Valvi, A.C., Fragkos, K.C., 2012. Critical review of the e-loyalty literature: a purchase-centred framework. Electr. Commerce Res. 12 (3), 331–378.

Venkatesh, V., Davis, F.D., 2000. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage. Sci. 46 (2), 186–204.
Weathers, D., Sharma, S., Wood, S.L., 2007. Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. J. Retailing 83 (4), 393–401.
Weiss, A.M., Lurie, N.H., MacInnis, D.J., 2008. Listening to strangers: whose responses are valuable, how valuable are they, and why? J. Mark. Res. 45 (4), 425–436.
Yue, W., 2015. The list of China's top 100 cities with highest GDP in 2015 (in Chinese). Available at: hn.rednet.cn/c/2016/01/23/3895505.htm.