Provider Feedback Information and Customer Choice Decisions on Crowdsourcing Marketplaces: Evidence from Two Discrete Choice Experiments

This is an author pre-print version of a paper forthcoming in Decision Support Systems


© 2015. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

http://creativecommons.org/licenses/by-nc-nd/4.0/
Provider Feedback Information and Customer Choice Decisions on Crowdsourcing Marketplaces: Evidence from Two Discrete Choice Experiments

Behrang Assemiab1, Daniel Schlagweina
a UNSW Australia Business School, Australia.
b Faculty of Engineering, Architecture and Information Technology, University of Queensland, Australia.

Abstract

Crowdsourcing marketplaces are increasingly becoming popular for the online transactions of services. On these marketplaces, profile information of providers, especially feedback left by previous customers, is the main information source for choice decisions of prospective customers. In the study reported in this paper, we examined the impacts of various feedback information components on provider profiles on the decisions of customers. We conducted two fractional factorial discrete choice experiments, one in a controlled laboratory setting and one online on a crowdsourcing marketplace. We found that the feedback information components “number of reviews” and “average weighted rating” have the largest impacts on the decisions of customers. We also found that “positive ratings” and “positive comments” have significant impacts on customers’ decision-making, especially when they appear on the first feedback page. We also found in the lack of highly visible feedback components on the subsequent feedback pages, “negative comments” become a significant determinant of customers’ decisions. We also showed the significant impact of information consistency on customers’ decision-making, through the synergistic interaction effects between different feedback components. Finally, we found some evidence that the cost of evaluating a feedback information component has a negative impact on the likelihood of customers evaluating that

1 Corresponding author

Email addresses and phone numbers: b.assemi@unsw.edu.au, +61 7 3365 3520 (B. Assemi); schlagwein@unsw.edu.au, +61 2 9385 6487 (D. Schlagwein)

Abbreviations: DCE (Discrete Choice Experiment), AMT (Amazon Mechanical Turk).
information component. The article concludes with implications of the findings of the study for theory and practice.

**Keywords:** Crowdsourcing, provider profile, feedback information, customer decision, choice decision, experimental research, discrete choice experiment, nested logit model.

1 **Introduction**

In the past few years, online crowdsourcing marketplaces (also often referred to as freelancing marketplaces and electronic service marketplaces) have become of central importance for outsourcing of services [1-3]. Crowdsourcing marketplaces allow customers to outsource their service needs to providers, typically using reverse auctions or private negotiations to determine prices [2, 4]. The crowdsourcing marketplaces industry in the US is expected to grow at a rate of 35% per year in the decade 2010-19 [5].

Deciding on the most appropriate provider is a key challenge for customers on crowdsourcing marketplaces [3]. Customers base their decisions primarily on provider profile information [3, 6-8]. They face two main problems in their decision-making. Firstly, the information on the profile may be incomplete. That is, not all information components relevant for customer decisions are available on provider profiles [6]. There is an “information asymmetry” between providers and customers (i.e., providers know their service quality better than customers can know) [6, 9]. Secondly, the information on the profile may be irrelevant for customer decisions, which amplifies information overload [10]. Given the bounded rationality of customers (e.g., limited time available to make a decision), they cannot effectively evaluate all the information components available on provider profiles, both due to the large number of information components and the large number of competing providers [10-12].

These problems lead to crowdsourcing marketplaces being characterized by high levels of uncertainty and low levels of trust [2, 3, 13]. Providers are incentivized to overstate their qualities (i.e., “moral hazard”), and customers are often unable to make good decisions (i.e.,
“adverse selection”) [14, 15]. Crowdsourcing marketplace provider profiles do not sufficiently help customers to distinguish between low- and high-quality providers [16]. Hence, customers are usually more willing to transact with their previous providers, if they have an acceptable experience, instead of searching for new, better providers [3, 8].

The usefulness of provider profiles on crowdsourcing marketplaces can be improved through design changes by the crowdsourcing marketplaces. To improve provider profiles’ design, however, we need to better understand how different information components on these profiles affect customers’ decision-making. At present, our knowledge about the relevance and importance of the information components that customers consider when selecting a crowdsourcing marketplace provider is fragmented [6]. Prior research suggests that customer decisions are primarily driven by the information components that reflect previous customers’ feedback, due to the higher credibility of these information components [17]. While studies found a positive impact of some feedback information components (e.g., “average rating”) on customers’ decision-making, the impact of other feedback components (e.g., “negative comments” hidden in not immediately visible feedback pages), their relative importance and their potential interactions are not considered well [17]. Hence, our study is set to improve our knowledge of such impacts by answering the following research question: How do the different provider profiles’ feedback information components impact on customer choice decisions on crowdsourcing marketplaces?

Taking advantage of recent methodological progress in discrete outcome modeling, we designed a discrete choice experiment (DCE) and ran it in two different settings: a controlled university laboratory and a crowdsourcing marketplace. We used nested logit analysis to evaluate a set of proposed hypotheses based on the data collected in the two experiments. The main benefit of DCEs is that they more closely resemble people’s real-life decision-making when selecting the best alternative (in comparison with the other methods of evaluating customer preferences) [18]. Nested logit modeling also provides valuable, detailed insights
about people’s decision process, overcoming deficiencies of earlier analysis techniques (e.g., multinomial logistic regression) [19].

The remainder of this article is structured as follows. In section 2, we review the existing literature on online profile information. In section 3, we develop a theoretical model informed by this review. In section 4, we present our research method. In section 5, we present the empirical findings of our study. In section 6, we discuss the theoretical meaning of these findings. In section 7, we conclude the article with implications for scholars and practitioners.

2 Literature Review

This research builds on, and contributes to, the literature interested in the impacts of profile information in the context of crowdsourcing marketplaces and similar IT-enabled online exchanges for services (we also consider, with caution, the related literature on electronic marketplaces for products [e.g., Amazon], which have more standardized items and hence less quality uncertainty). The literature has established that online feedback mechanisms are effective means to build and represent the reputation and trustworthiness of providers [17, 20-23]. Feedback information on provider profiles is more important than other information components, such as self-descriptions of providers. This is so because feedback information components reflect the genuine and de facto experiences of past customers [12]. Hence, all electronic marketplaces, including crowdsourcing marketplaces, encourage customers to leave feedback after each transaction with a provider [24, 25].

According to the literature, feedback information components reduce the information asymmetry between providers and customers in electronic marketplaces [16, 17, 25]. Hence, feedback information components help to develop trust on electronic marketplaces, including crowdsourcing marketplaces [16, 26]. The existence of feedback mechanisms prevents opportunistic behavior by an online provider because such behavior would become permanently visible on the provider’s profile [15], and damage the provider’s gradually established reputation [17, 27].
Which feedback information components are the most important? Thus far, the literature appears to agree that across contexts the “average rating” (average rating based on all past customer ratings) is a key information component. A high average rating positively affects the decisions of customers to transact with the corresponding provider [28, 29]. A high average rating also positively correlates with the likelihood of this provider actually being paid by their customers [21].

The literature also suggests that the “number of reviews” is relevant on online marketplaces [21, 23, 29]. While these findings are related to the transaction behavior of customers in the context of marketplaces for products, it appears reasonable to assume that the same underlying logic (that customers are more likely to select providers that have been selected more frequently by previous customers) also may apply to crowdsourcing marketplaces.

Other provider feedback components may be also important for customer decision-making. While not investigated in the context of crowdsourcing marketplaces, the literature on marketplaces for products found that the number of “positive ratings” and the number of “positive comments” can affect customers’ transacting behavior [17, 27].

Based on our review, where we are lacking knowledge in both the literature on crowdsourcing marketplaces in specific, and the literature on online marketplaces in general are: a) we do not clearly know the impact of feedback information components on customers’ choice decisions, as previous studies have evaluated the impact of these components on customers’ trust, price premiums, and bid prices rather than choices [27, 30]; b) we do not know the role of implicit characteristics of feedback information components in customer decision-making, for example, we do not know to what extent the visibility of information components matters; and c) we do not know how the information components matter in their relative relation to one another, for example, are there dominant effects? Are there interaction effects? We cannot answer these questions based on the existing literature. Table 1 summarizes the findings and gaps of the existing literature on the impacts of online feedback information on marketplaces.
Table 1: Existing Findings on Provider Feedback Information Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Studies</th>
<th>Findings</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reviews</td>
<td>Kim [29], Gefen and Carmel [21], Duan, Gu and Whinston [23]</td>
<td>A higher number of feedback reviews are positively associated with a higher likelihood that the respective provider is selected by customers.</td>
<td>Impact of number of reviews in relation to other information components has not been investigated (i.e., relative strength of effect and interaction effects).</td>
</tr>
<tr>
<td>Average (weighted) rating</td>
<td>Kim [29], Gefen and Carmel [21], Qu, Zhang and Li [22], Hong and Pavlou [16], Banker and Hwang [7]</td>
<td>High average rating on a provider profile positively impacts on the decisions of customers to transact with the provider.</td>
<td>Crowdsourcing marketplaces usually present an average weighted rating on each profile instead of a simple arithmetic average rating like other electronic marketplaces. In fact, the average weighted rating does not simply reflect the values of individual ratings, as the average is weighted by the respective project values, which are hidden from customers. The impact of this component, especially in relation to other components has not been investigated.</td>
</tr>
<tr>
<td>Positive ratings (absolute number or relative number in relation to negative/neutral ratings)</td>
<td>Duan, Gu and Whinston [23], Pavlou and Dimoka [17], Ba and Pavlou [27]</td>
<td>The total number of positive ratings is positively associated with the perceived trustworthiness of the respective provider.</td>
<td>Impact of number of positive ratings in relation to other feedback components has not been investigated (i.e., relative strength of effect and interaction effects). Not clear if effect of absolute number changes based on the presence of negative ratings (i.e., are there different effects?). Ratings are often displayed on multiple feedback pages – the role of their visibility is unknown.</td>
</tr>
<tr>
<td>Positive and negative comments (absolute number or relative number)</td>
<td>Li and Hitt [31], Dellarocas [25], Lee, Park and Han [32], Pavlou and Dimoka [17], Lee and Koo [26]</td>
<td>Individual positive or negative comments are positively associated with product sales. The temporal attribute of the comments has direct impact on their effectiveness. The credibility of negative comments is higher than the credibility of positive comments.</td>
<td>Impact of number of positive or negative comments in relation to other information components has not been investigated (i.e., relative strength of effect and interaction effects). Not clear if effect of the absolute number of positive comments changes based on the presence of negative comments (i.e., are there different effects?). Comments are often displayed on multiple feedback pages – the role of their visibility is unknown.</td>
</tr>
</tbody>
</table>

3 Theoretical Model

Building on the empirical findings of the literature, we have developed a theoretical model of the impact of provider feedback information on customers’ choice decisions on crowdsourcing marketplaces. We have focused on eight feedback information components on provider profiles that we found to be common on provider profiles across the leading crowdsourcing
marketplaces (i.e., Freelancer.com, UpWork [previously eLance and oDesk] and Guru; as of 2015).

In addition to the empirical findings, we also found signaling theory [33, 34] to help us theorize the “how” and “why” of effects of crowdsourcing marketplace provider profile information component. The core idea of signaling theory is that “signals” transmitted from a “source” to “receivers” convey information regarding unobserved qualities of the signaling source [34]. Other research in the area of online platforms has found this theory helpful as well [15, 16]. According to the theory, the information components of a provider profile signal the reputation and abilities of the respective provider (unobserved qualities of the source) to prospective customers (i.e., receivers) [17, 25]. Such signals help customers to distinguish between low- and high-quality providers [15, 24, 35].

Seen as a signal, there is consensus in the literature that visibility (how visible a signal is) and cost (how difficult to generate the signal is) of a feedback information component are most important for its credibility [14, 16, 34, 36]. That is, on crowdsourcing marketplace provider profiles, a high “number of feedback reviews” as well as a high “average weighted rating” are the most visible feedback information components. They are placed (on the above platforms, as of 2015) centrally, next to the name of the provider. They are also costly to generate signals: they can only be obtained by successfully completing projects for many clients over time [17, 25]. The number of feedback reviews and the average weighted rating signal the “lifelong” reputation of the provider [17, 27]. Hence, we expect that:

**H1. A high number of feedback reviews on a provider’s profile have a positive impact on the customers’ decisions to choose the provider.**

**H2. A high average weighted rating on a provider’s profile has a positive impact on the customers’ decisions to choose the provider.**
Signals of the performance of a provider in individual jobs (e.g., individual ratings) are both less visible (not shown as prominently) and less costly (it requires more effort to perform well over time than to perform well in a single project). That is, in contrast to other information components, only the number of feedback reviews and the average weighted rating are highly visible and very costly signals of the overall long-term behavior of the provider. Hence, we hypothesize:

**H3.** The number of feedback reviews and the average weighted rating on a provider’s profile have a larger impact on the customers’ decisions to choose the provider than other feedback information components.

In addition to aggregated information, the literature suggests that individual ratings and comments on provider profiles may also impact on customer decisions. Individual ratings and comments on a provider profile are usually divided across multiple feedback pages. Based on the above arguments about signal visibility, we expect the impact of individual ratings and comments to be different, based on the location of these ratings and comments (i.e., because of their different visibility). Studies of marketplaces for products have also suggested that customers mainly evaluate the feedback ratings and comments that appear on the first feedback page [17, 25]. Furthermore, a higher frequency of an information component leads to a higher perceived credibility of the respective component [34]. Thus, we hypothesize that the frequency and visibility of positive ratings, positive comments, and negative comments on each feedback page can affect the decision-making of customers:

**H4.** a) A high number of individual positive ratings have a positive impact on the customers’ decisions to choose the provider; b) the impact is larger for positive ratings on the first feedback page than for positive ratings on subsequent feedback pages.

**H5.** a) A high number of individual positive comments have a positive impact on the customers’ decisions to choose the provider; b) the impact is larger for positive comments on the first feedback page than for positive comments on subsequent feedback pages.
H6. a) A low number of individual negative comments have a positive impact on the customers' decisions to choose the provider; b) the impact is larger for negative comments on the first feedback page than for negative comments on subsequent feedback pages.

The literature further suggests that individual text comments impact more strongly on provider's trustworthiness than individual numeric ratings [17, 27]. In fact, text comments provide “stronger” (i.e., information richer) signals than numerical ratings. Text comments (e.g., through a narrative explaining a positive or negative experience) can signal information beyond what is communicated through numerical ratings [17]. We put this argument to the test by focusing on positive ratings and comments on the first feedback page. Hence, we propose:

H7. Individual positive textual comments (on the first feedback page) on a provider's profile have a larger impact on customers' decisions to choose the provider than individual positive numeric ratings (on the first feedback page).

A final aspect that we consider important to understand is consistency between information components. Consistency across the contents of multiple feedback information components (i.e., consistency across signals) increases the perceived credibility of the respective feedback components [37]. For example, customers might be suspicious, if the average weighted rating is high, but based only on very few ratings. If this argument holds, customers would prefer providers with consistent feedback information above those with inconsistent feedback information (in addition to the impacts of individual components). We put this argument to the test, by studying the interaction effects between a high number of feedback reviews and a high average weighted rating, a high number of positive comments, as well as a low number of negative comments. Accordingly, we propose the following hypotheses:

H8. There is a positive interaction effect between the individual feedback components a) high average weighted rating, b) high number of positive comments [on the first feedback page],
and c) low number of negative comments [on the first feedback page], and high number of feedback reviews on a provider’s profile, on customers’ decisions to choose the provider.

4 Research Method

4.1 Discrete choice experiments

Discrete choice experiments (DCEs) suited our research objectives best; and we used DCEs to evaluate the above proposed set of hypotheses. DCEs are useful to predict how people’s choices will change with changes in the attributes of available alternatives [18, 40, 41]. These attributes usually entail conflicts (e.g., benefit vs. cost) and, thus, selecting among alternatives with particular attribute combinations implies trade-offs [40]. DCEs assume that individuals assign a utility to each attribute of the existing alternatives when selecting the best alternative. Eventually, in line with economic theory, individuals will choose the alternative with the maximum overall utility [39, 40]. Random utility models, such as multinomial logit [41], and its advanced variations such as nested logit and mixed logit, are then used to infer people’s preferences regarding attribute levels [19, 39, 40].

DCEs have two main advantages in comparison with other (conjoint analysis) methods in studying customer preferences [18]: Firstly, DCEs most closely resemble people’s actual decision-making when selecting the best alternative in real transactions (including the decision not to select any available alternative), because people’s choice represents the dependent variable. Secondly, DCEs allow the evaluation of the impact of other alternatives on the attractiveness of any specific alternative.

In DCEs, choice alternatives (in our case, provider profiles) are divided across several systematically formed choice sets [18]. To accurately reveal the impact of trade-offs underlying

---

2 DCEs (and corresponding random utility models used for the analysis of their results) gained their creator, Daniel McFadden, the 2000 Nobel Prize in Economics because they dramatically improve our ability to understand choice behaviors based on individuals’ preferences [38, 39]. To our knowledge, this is the first study in the information systems (IS) field to use this method.
the decision-making of participants, Pareto optimal choice sets\textsuperscript{3} are used. A choice set is Pareto optimal if none of the alternatives dominates, or is dominated by the rest of alternatives in that set (in our case, a set of provider profiles to choose from) [18, 40, 42, 43]. This is to avoid having participants make comparisons that cannot reveal the impact of the profile attributes, because of the trivial choice outcome.

Furthermore, DCEs typically include the “no choice” option in each choice set to make the decision-making of participants more realistic (in our case, the individual might choose not to crowdsource at all given his/her options), and to make the aggregation and analysis of the choices across choice sets possible [18, 44]. While conducting the experiment, participants (individuals) are presented some of the choice sets (in our case, several sets of provider profiles). The participants are then asked to evaluate the alternatives and choose an alternative they consider to be the best option (separately for each choice set) [18, 44]. The choices made by all participants in all choice sets are then aggregated with the help of the no choice option, and used for analysis [18].

4.2 DCE design

For this paper, we conducted identical-designed DCEs in two different contexts. We conducted one DCE (n1 = 120) in a controlled laboratory environment at a research-intense university, using online recruitment system ORSEE [45]. We conducted a second, replication DCE (n2 = 695) in a real-world crowdsourcing marketplace with participants experienced in using crowdsourcing marketplaces as both providers and customers, using Amazon Mechanical Turk (AMT) for recruitment (on the suitability of AMT for experimental research see [46-50]).

Furthermore, prior to the DCEs, we had conducted a survey on the crowdsourcing marketplace to accurately design our DCE, especially to find out what constitute “low” and “high” levels of

\textsuperscript{3} A choice set is called “Pareto optimal”, when for every two different alternatives \((x_1, x_2, \ldots, x_m)\) and \((y_1, y_2, \ldots, y_m)\) in the choice set \((x_i \text{ or } y_i \text{ is the } i\text{th attribute of an alternative at one of its possible levels}),\) subscripts \(i\) and \(j\) \((i \neq j)\) can be found that \(x_i < y_i\) and \(x_j > y_j\). That is, Pareto-optimal configuration prevents dominant or dominated alternatives in the choice set [18, 40].
different profile attributes to participants (i.e., we determined these levels empirically, see below for further details). Finally, we conducted a post-DCE survey (immediately after the DCE) with the participants of the lab experiment to gain further quantitative (e.g., self-explicated perceived importance of different profile attributes) and qualitative data (e.g., free-text comments) on their decision processes.

In the design used in both DCEs, we created 64 \(2^8 - 2\) crowdsourcing provider profiles, following a resolution V fractional factorial design \([51]\). The resolution V fractional factorial design that we applied allowed us to reduce the number of profiles to a quarter, while still being able to evaluate all main effects and two-way interactions between attributes \([40, 51]\).

We excluded two crowdsourcing provider profiles (one was always dominating others, one was always dominated by others), and assigned the remaining 62 crowdsourcing provider profiles to 17 Pareto optimal choice sets. We used the following formula to construct six Pareto optimal choice sets:

\[
S_i = \{(x_1, x_2, ..., x_m) | \sum_{i=1}^{m} x_i = l \}, l = 0, 1, ..., \sum (s_i - 1)
\]

For a DCE with \(m\) attributes (each attribute \(i\) at \(s_i\) levels represented by 0, 1, ..., \(s_i - 1\), for \(i = 1, 2, ..., m\)), the sum of the attribute levels in each profile \((x_1, x_2, ..., x_m)\) is between 0, and \(\sum (s_i - 1)\). If the attribute levels are sorted from lower costs to higher benefits, choice sets \(S_i\) in the formula are the Pareto optimal subsets of \(S\), where \(S\) is the set of all profiles \([18, 52]\). Raghavarao and Wiley \([40]\) showed that in an \(s^m\) experiment, all \(m(m - 1)(s - 1)^2/2\) contrasts of two-way interaction effects can be estimated by using Pareto optimal choice sets \(S_i, S_{i+1}\), and \(S_{i+2}\), where \(2(s - 2) \leq l \leq (m - 2)(s - 1)\). Thus, we divided the 62 profiles of our

---

\(\text{We used} \ 8-2.1 \text{ block of the design matrix with the defining relation of } I = ABCDEG = ABCFH = DEFGH \) (with \(A\) being the first attribute, \(B\) being the second attribute, etc.). This design ensures that the main effects and the two-way interactions are all orthogonal to the block and the minimum possible aberration affects them \([51]\).
DCE across the six possible Pareto optimal choice sets of $S_1$ to $S_6$ to estimate the main effects and all two-way interactions ($s = 2$ and $m = 8$ in our experiment).

We finally divided each of these six Pareto optimal choice sets into smaller sets to make it easier for research participants to compare the provider profiles in each choice set (any subset of a Pareto optimal choice set is also a Pareto optimal choice set) [18, 40]. The resulting choice sets each consisted of three or four profiles. For the lab experiment, we divided the choice sets between four experimental sessions by randomly assigning four or five choice sets to a session. In each session, we asked the participants to evaluate each choice set separately. The participants then could choose their preferred provider profile within each choice set. We included the “no choice” option in each choice set, as discussed above.

Based on the theoretical hypotheses proposed in section 3, we created idealized crowdsourcing marketplace provider profiles for our DCEs (an example of a set of four profiles is shown in Figure 1). In each profile, we included the eight attributes of theoretical interest (i.e., information components), which enabled us to evaluate the impact of these attributes in crowdsourcing marketplaces. To determine the impact of each attribute on the decision-making of participants, we considered two perfectly distinct, quantitative levels for each attribute (i.e., particular levels of low/high) as suggested by Hair, Anderson, Black, Babin and Black [53].

![Figure 1. Sample choice set and profiles](image)
Table 2 shows these eight attributes, and their low and high-level values. We used the findings of the broader literature on online feedback reviews [e.g., 21, 22, 25, 27] as a guide for the levels in Table 2. However, we conducted a survey (with individuals of the same profile as our eventual DCE participants) to determine meaningful attribute levels.

Table 2: Attribute levels for provider profile information components

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of feedback reviews</td>
<td>50</td>
<td>375</td>
</tr>
<tr>
<td>Average weighted rating</td>
<td>2</td>
<td>4.8</td>
</tr>
<tr>
<td>Number of positive ratings on the first feedback page</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>Number of positive comments on the first feedback page</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Number of negative comments on the first feedback page</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Number of positive ratings on each subsequent feedback page</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>Number of positive comments on each subsequent feedback page</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Number of negative comments on each subsequent feedback page</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

As a best practice in previous experimental studies [e.g., 54], and recommended by Eriksson, Johansson, Kettaneh-Wold, Wikstrom and Wold [55], we used a range of values around the determined low/high values in the experimental profiles to make them more realistic. For example, for ratings, the range of 1.8–2.0 was used to present low ratings, and the range of 4.7–5.0 was used to present high (i.e., positive) ratings.

While the number of positive and negative text comments on each feedback page of provider profiles was determined on the same way as the other six attributes (by the design matrix and the respective attribute levels), the contents of these text comments were randomly extracted for each research participant from an existing repository of 60 comments.

The applied attribute levels were not correlated. In our experimental profiles, the average weighted rating and individual numerical ratings of a provider are the most likely to be

---

5 For the online survey to determine the attribute levels, we recruited 30 different participants on AMT. We restricted the participants to workers who had more than 1,000 approved jobs, and at least a 95% approval rate for their previous jobs. Four cases with inconsistent response values were omitted from the response set. Thus, we used the remaining 26 responses to specify the attribute levels. The overwhelming majority of the participants (96.1%) had online marketplace experiences. Considering the minimum, maximum, median and mean of the values specified by the participants for each attribute level, we determined the attribute levels shown in Table 2.

6 We created this repository which contains 41 very positive text comments and 19 very negative text comments, relying on the categorization of text comments at eBay.com, as proposed by Pavlou and Dimoka [17]. We refined the comments originally extracted from real crowdsourcing marketplace profiles to replace provider identities with one of the general terms “provider” or “supplier”. Each comment contains between 200 and 500 words.
correlated theoretically. However, the average weighted rating is not a simple arithmetic mean of all individual ratings assigned to the provider. It is the average job rating of the provider weighted by his/her earnings per job. Provider earnings are not necessarily illustrated for each job on provider profiles and, accordingly, we have not included them in our experiment. Thus, a provider may possess a high average weighted rating even though the majority of his/her individual ratings are low. This means that the provider has accomplished some high value projects for which he/she has obtained good ratings. Such explanations were included in the profile descriptions to avoid confusion among participants.

4.3 Procedures

Owing to the highly customized design of profiles that we used in order to make them similar to real crowdsourcing marketplace provider profiles, we developed our own web application to conduct the experiments (Figure 1 and Figure 2 are screenshots of the application). The choice sets, consisting of the generated provider profiles based on different levels of the attributes, were hosted in this web application. The web application also handled participants’ navigation and data entry.

For the controlled laboratory experiment, we recruited 120 participants. The choice sets were randomly divided across four experimental sessions, and the participants were randomly assigned to these four experimental sessions, with 30 per session. Each participant evaluated four or five choice sets. Thus, each of the 17 choice sets was evaluated by 30 participants. Every participant was paid AU$22 for participating in an experimental session. Each session lasted approximately 50 minutes.

For the online experiment on the crowdsourcing marketplace, we recruited 695 participants. Again, the choice sets were randomly assigned to the participants, while each set was evaluated

---

7 We used Microsoft ASP.Net 2008 and Microsoft SQL Server 2005 to develop the web application. We deployed our web application on a well-known Windows host with 99.9% uptime guarantee. We replicated the web application on another Windows host to be used as the alternative in case of any potential failures (although it was never used). We tested the application with 10 users and improved some parts of it, based on the feedback we received from these users.
by a minimum of 39 to a maximum of 43 participants. Every participant was paid AU$0.50 per evaluation (each evaluation took about 7 minutes on average).

In the experiments, our web application briefed the participants with a scenario in which they considered outsourcing a service typically traded on crowdsourcing marketplaces (i.e., a web development project, as shown in Figure 2). Next, our web application showed choice sets of crowdsourcing marketplace provider profiles (as shown in Figure 1) to the participants. In each set, the profiles were illustrated side by side in separate frames, while the participants were also able to see each profile in the full window mode. The participants had at least 10 minutes to compare the profiles in each set, allowing them to make an informed decision about the best provider in the set, without any time pressure as emphasized in the literature [56]. Each participant then nominated his or her preferred provider within each choice set (or decided not to crowdsourc at all).

![Figure 2. Screenshot of web project and profiles description](image)

We report the results of our DCEs in section 5. However, first we need to discuss how we analyze the choices.
4.4 Data analysis via discrete choice random utility models

We used a multinomial logit random utility model [41] to analyze the participants’ choices in our DCEs. Probabilistic choice models (such as multinomial logit) are the most common techniques used to analyze the results of DCEs [18, 57]. The widespread application of these techniques is mainly due to their capabilities in providing the probabilities of each alternative, based on the values of independent variables. These techniques presume that the probability of selecting each alternative is directly related to the attributes characterizing that alternative [18]. Therefore, these probabilities are used to estimate the impact of each attribute on the choice of alternatives [58].

Based on utility theory, the probability of choosing profile \(i\) for observation \(n\) \((i \in I, \text{where } I\) specifies all alternatives in observation \(n\)), \(P_n(i)\), is [18, 19]:

\[
P_n(i) = \frac{e^{V_i}}{\sum_{\forall \theta} e^{V_\theta}}
\]

in which, \(V_i\) is the utility of profile \(i\), estimated by the profile attributes.

Applying the multinomial logit modeling, \(T_{i\in}\), a linear function which determines the choice of profile \(i\) for observation \(n\), is considered as \(T_{i\in} = \beta_iX_{i\in} + \epsilon_{i\in}\), where \(\beta_i\) is a vector of estimable parameters for profile \(i\), \(X_{i\in}\) is a vector of the attributes that determine choices for observation \(n\), and \(\epsilon_{i\in}\) is the disturbance term. Thus,

\[
P_n(i) = P(T_{i\in} \geq T_m) = P(\beta_iX_{i\in} + \epsilon_{i\in} \geq \beta_jX_{j\in} + \epsilon_{j\in}), \quad \forall j \neq i
\]

Assuming a random extreme value distribution [59] for the disturbance term, while all \(\epsilon_{i\in}\) are independently and identically distributed, the previous formula can be revised to [19]:

\[
P_n(i) = P \left( \beta_iX_{i\in} + \epsilon_{i\in} \geq \max_{\forall \theta \neq i} (\beta_iX_{\theta\in} + \epsilon_{\theta\in}) \right) = \frac{\text{EXP}[\beta_iX_{i\in}]}{\sum_{\forall \theta}[\beta_iX_{\theta\in}]}
\]
which is the standard multinomial logit formulation. The model is estimated by maximizing
the log likelihood of the alternatives [18]. In fact, the results of the model estimation indicate
the likelihood of the selection of each alternative, based on the value of attributes $X_i$. The
following log likelihood function is applied to estimate the parameters (i.e., $\beta$’s):

$$LL = \sum_{n=1}^{N} \left( \sum_{i=1}^{I} \delta_{in} \left[ \beta_i X_{in} - \ln \sum_{v_l} \text{EXP}(\beta_i X_{in}) \right] \right)$$

where $\delta_{in}$ is 1, if the profile $i$ is chosen for observation $n$, and zero otherwise.

However, the assumption of independence of disturbance terms ($\epsilon_{in}$’s) among alternatives,
which is known as the independence of irrelevant alternatives (IIA), is very limiting. The nested
logit model is a widely used modeling alternative for the standard multinomial logit that
overcomes the IIA limitation [19]. The model groups different alternatives that potentially
share unobserved effects into nests to address the issue of correlations among their
disturbance terms [19]. However, nesting is not equivalent to a decision tree or a
representation of hierarchical decision-making process [19]. To illustrate, Figure 3 shows an
example of the nested structure of the DCE profiles used for the analysis of the lab data.

![Figure 3: Example nested structure of profile choice](image)

Applying the nested logit model, McFadden (1981) has shown that the probability of choosing
alternative $i$ in observation $n$ can be estimated using the following formulas, given the
disturbance terms are extreme value distributed:

$$P_n(i) = \frac{\text{EXP} [\beta_i X_{in} + \Phi_i LS_{in}]}{\sum_{v_l} \text{EXP} [\beta_i X_{in} + \Phi_i LS_{in}]}$$
\[ P_n(j|i) = \frac{\exp[\beta_{ji}X_n]}{\sum_{xj} \exp[\beta_{ji}X_n]} \]

\[ LS_{in} = \ln \left( \sum_{xj} \exp[\beta_{ji}X_n] \right) \]

where \( P_n(i) \) is the unconditional probability of choosing profile \( i \) for observation \( n \), \( P_n(j|i) \) is the probability of choosing profile \( j \) for observation \( n \) conditioned on the profile being in the nest of profiles \( i \), \( J \) is the conditional set of profiles (conditioned on \( i \)), \( I \) is the unconditional set of nests, \( LS_{in} \) is the inclusive value (logsum) and \( \Phi_i \) is an estimable parameter. According to McFadden (1981), \( \Phi_i \)'s should be greater than 0 and less than 1 to conform to the model derivation. The difference between each \( \Phi_i \) and 1 shows the significance of the assumed shared unobserved effects in the corresponding nest [19]. As recommended by the literature, we used advanced software packages to estimate all nests simultaneously, using full-information maximum likelihood [19]. In particular, we used R [60] and NLogit 5.0 [61] to conduct the analysis.

5 Analysis Results

The university lab sample (first DCE) was a typical student population in demographic terms. The participants were between 18 and 38 years old: 61\% of the participants were female, and 39\% were male. In all, 34\% had graduate degrees, 30\% had bachelor's degrees, and 36\% did not have a university degree (yet). 94\% had experience on electronic marketplace, with 6\% not reporting such experience.

The crowdsourcing marketplace sample (second DCE) consists of a broader range of participants. The participants were between 20 and 62 years old: 45\% of the participants were female, and 55\% were male. In all, 31\% had graduate degrees, 53\% had bachelor's degrees, and 16\% did not have a university degree. As active marketplace participants, all of the participants had (different levels of) first-hand crowdsourcing marketplace experience.
Table 3: Summary of nested logit analysis results and self-explicated ratings

<table>
<thead>
<tr>
<th>Hypothesis and Parameters</th>
<th>University Lab Sample</th>
<th></th>
<th></th>
<th>Crowdsourcing Marketplace Sample</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-Value</td>
<td>Max Elasticity</td>
<td>Explicated Ratings</td>
<td>Coefficient</td>
<td>t-Value</td>
</tr>
<tr>
<td>H1 and H3: No. of feedback reviews</td>
<td>0.508***</td>
<td>2.65</td>
<td>0.050</td>
<td>17.75</td>
<td>0.936***</td>
<td>7.08</td>
</tr>
<tr>
<td>H2 and H3: Average weighted rating</td>
<td>1.650***</td>
<td>8.37</td>
<td>0.101</td>
<td>24.65</td>
<td>0.893***</td>
<td>7.00</td>
</tr>
<tr>
<td>H4a and H4b: No. of positive ratings on first feedback page</td>
<td>0.081</td>
<td>0.43</td>
<td>0.008</td>
<td>10.45</td>
<td>0.204*</td>
<td>1.82</td>
</tr>
<tr>
<td>No. of positive ratings on each subsequent feedback page</td>
<td>0.147</td>
<td>0.73</td>
<td>0.015</td>
<td>6.11</td>
<td>-0.245</td>
<td>-1.16</td>
</tr>
<tr>
<td>H5a, H5b and H7: No. of positive comments on first feedback page</td>
<td>1.116***</td>
<td>5.73</td>
<td>0.093</td>
<td>13.08</td>
<td>0.291***</td>
<td>2.81</td>
</tr>
<tr>
<td>No. of positive comments on each subsequent feedback page</td>
<td>0.891**</td>
<td>2.02</td>
<td>0.048</td>
<td>7.35</td>
<td>0.707**</td>
<td>2.20</td>
</tr>
<tr>
<td>H6a and H6b: No. of negative comments on first feedback page</td>
<td>0.063</td>
<td>0.33</td>
<td>0.006</td>
<td>12.82</td>
<td>-0.055</td>
<td>-0.45</td>
</tr>
<tr>
<td>No. of negative comments on each subsequent feedback page</td>
<td>-0.304</td>
<td>-0.65</td>
<td>-0.019</td>
<td>7.79</td>
<td>0.696**</td>
<td>2.30</td>
</tr>
<tr>
<td>H8a: No. of feedback reviews × Average weighted rating</td>
<td>0.224***</td>
<td>4.40</td>
<td>0.021</td>
<td>N/A</td>
<td>0.090**</td>
<td>2.12</td>
</tr>
<tr>
<td>H8b: No. of feedback reviews × No. of positive comments on first feedback page</td>
<td>-0.100**</td>
<td>-2.15</td>
<td>-0.013</td>
<td>N/A</td>
<td>0.064*</td>
<td>1.94</td>
</tr>
<tr>
<td>H8c: No. of feedback reviews × No. of negative comments on first feedback page</td>
<td>0.036</td>
<td>0.63</td>
<td>0.006</td>
<td>N/A</td>
<td>-0.009</td>
<td>-0.25</td>
</tr>
<tr>
<td>“No choice” intercept</td>
<td>0.611</td>
<td>1.63</td>
<td>N/A</td>
<td>N/A</td>
<td>0.620**</td>
<td>2.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Branch</th>
<th>IV Parameter</th>
<th>Std. Error</th>
<th>t-Value of 1-IV</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Branch 1</td>
<td>0.886</td>
<td>0.167</td>
<td>0.68</td>
<td>N/A</td>
<td>0.535</td>
</tr>
<tr>
<td>Branch 2</td>
<td>0.548</td>
<td>0.114</td>
<td>3.96</td>
<td>&lt;0.001</td>
<td>N/A</td>
</tr>
<tr>
<td>“No choice” branch</td>
<td>1.0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* Not significant, **p < 0.1, ***p < 0.05, ****p < 0.01, Number of negative comments is reverse coded.

Branches for the lab sample: 1) “High average weighted rating”, and 2) “Low average weighted rating”
Branches for the online sample: 1) “High number of positive ratings on each subsequent feedback page”, and 2) “Low number of positive ratings on each subsequent feedback page”

The results of the nested logit estimation for both samples, and self-explicated importance ratings for the lab sample are summarized in Table 3. The overall goodness-of-fits of the
models for both samples were satisfactory [19, 40]: log likelihood ratio $LR = -598.44$, $\chi^2_{14} = 3069.24$ and $p < 0.001$ for the university lab sample (first DCE), and log likelihood ratio $LR = -975.15$, $\chi^2_{14} = 3251.14$ and $p < 0.001$ for the crowdsourcing marketplace sample (second DCE). The effect sizes of the models were also satisfactory, given the adjusted McFadden pseudo R-squared$^8 [62]$ of $R^2 = 0.717$ for the first model, and $R^2 = 0.623$ for the second model. Maximum elasticities of the attributes are shown in the table, as we are interested in comparing the largest possible effect size of these attributes.

As shown in the lower part of Table 3, the nested logit models for both samples have three branches (i.e., branch 1, branch 2, “no choice” branch). Nesting the “no choice” option in a separate branch is trivial because of its totally different nature. However, the experimental profiles were nested into two branches based on the value of their “average weighted ratings” for the lab sample, because this value could cause the participants to evaluate the respective profiles differently. Given the relatively high number of profiles that were evaluated by the lab participants (i.e., 14 to 18 profiles), the highly visible average weighted rating component could guide the participants in more efficiently evaluating the profiles. Our reason for this nesting is that the participants evaluate individual ratings and comments more carefully, only when the average weighted rating is low (because the time cost of such more detailed evaluation is high). The results also show that the participants’ evaluation of the profiles with low average weighted rating is significantly different from other profiles.

Similarly, in addition to the no choice branch, the model for the crowdsourcing marketplace sample includes two branches. However, the experimental profiles were nested into these two branches based on their “number of positive ratings on each subsequent feedback pages”. In contrast to the lab experiment, the crowdsourcing marketplace participants evaluated less profiles (i.e., 3 or 4 profiles). Thus, they would potentially evaluate all the information

$^8$ McFadden Pseudo $R^2 = \frac{-2LL(0)-(2LL(B))}{-2LL(0)}$, in which, $LL(B)$ is the log likelihood of the full model, and $LL(0)$ is the log likelihood of the null model.
components on the first feedback page. Furthermore, the number of positive ratings on each subsequent feedback page could guide them on how to evaluate the subsequent feedback pages of the profiles, as this component is the most visible on these pages. Our reason for this nesting is that the participants evaluate individual feedback comments on subsequent feedback pages more carefully, when the number of positive ratings on these pages is low (as the cost of such evaluation is high). The results also show that the participants’ evaluation of the profiles with low versus high number of positive ratings on each subsequent feedback page is significantly different from each other as well as the no choice option.

6 Discussions

We hypothesized that the higher number of feedback reviews on a provider’s profile would positively impact on the likelihood of customers choosing the provider (H1). As shown in Table 3, the impact of this component is significant in both samples. H1 is supported (i.e., not falsified) by the data of both samples.

Our theoretical explanation for the significant impact of the number of feedback reviews is that a higher number of feedback reviews indicates that the provider has been successful in persistently attracting customers to outsource their projects to the provider. For example, a provider with 1,000 feedback reviews is perceived as more trustworthy compared to a provider with the same average weighted rating, but only 10 feedback reviews. However, we note that the findings of another recent study of crowdsourcing marketplace provider profiles, namely Gefen and Carmel [3], do not suggest a significant impact of the number of feedback reviews on customers’ decision-making. Further research is needed to resolve or better understand these conflicting findings.

We hypothesized that a higher average weighted rating on a provider’s profile would be positively associated with the likelihood of customers choosing the provider (H2). As illustrated in Table 3, the impact of the average weighted rating on customers’ decision-making was positive and significant for both samples. H2 is supported by the data.
We hypothesized that the effects of the number of feedback reviews and the average weighted rating are stronger compared to the rest of the feedback information components (H3). As illustrated in Table 3, the maximum elasticity of the number of feedback reviews is 0.05 and 0.10 for the university lab sample and the crowdsourcing marketplace sample, respectively. The maximum elasticity of the average weighted rating is also 0.10 and 0.09 for the two samples, respectively. Given these elasticities and the corresponding attributes’ coefficients, the impact of the average weighted rating is much larger than the rest of the components for both samples, while the impact of the number of feedback reviews is larger compared to the rest of components only for the online sample. However, the stated preference of the participants obtained through the self-explicated ratings in the lab experiment shows that these two information components are the most important determinants of customers’ decisions in a crowdsourcing marketplace (as shown in Table 3). We attribute these large impacts to the highly visible and costly to generate nature of these two components. H3 is supported by the data.

One interesting finding is that the impact of the interaction between the number of feedback reviews and the average weighted rating is not only significant but actually larger than the impact of the number of feedback reviews for the university lab sample (as discussed below for H8a). Thus, the potential customers give more weight to the number of feedback reviews, when the providers have high average weighted rating. This is reasonable as a high number of feedback reviews with a low average weighted rating mean to the customer that the corresponding provider has somehow obtained many jobs, but performed badly in many of them.

Furthermore, we hypothesized that a high number of positive ratings, a high number of positive comments, and a low number of negative comments would be positively associated with the likelihood of customers choosing the provider (H4a, H5a, and H6a, respectively), and that their effects were larger on the first feedback page (H4b, H5b and H6b, respectively). The self-explicated ratings, as shown in Table 3, support these hypotheses. However, the revealed
preference of the participants obtained through the experiments implies a more complicated process of decision-making, as discussed next.

The results of the nested logit analysis for the online sample indicate that the number of positive ratings on the first feedback page has a significant positive impact on the participants’ decisions, while the impact of the number of positive ratings on each subsequent feedback pages is not significant. However, the analysis of the university lab data shows no significant impact of the number of positive ratings at all. This means that the lab participants do not assign a high weight to the numerical ratings, compared to feedback comments (as discussed below in relation to H7). Thus, H4a is supported by the crowdsourcing marketplace data, but not by the university lab data, while H4b is not supported at all.

The results of the analysis show that the impact of the number of positive comments is positive and significant on the first and subsequent feedback pages for both samples. Furthermore, as hypothesized, this impact is larger for the first feedback page compared to subsequent feedback pages (see the elasticities and coefficients of the number of positive comments on the first and each subsequent feedback pages for both samples). H5a and H5b are supported.

Another interesting finding is the very large impact of the number of positive comments on the first feedback page (given 0.09 as the corresponding maximum elasticity) compared to the rest of the feedback components for the university lab sample. This impact is even larger than the impact of the number of feedback reviews for the same sample (with the maximum elasticity of 0.05), although initially we expected to see a larger impact for the number of feedback reviews. This finding implies that the lab participants assign more weights to the contents of the feedback comments appearing on the first feedback page, rather than their overall count.

The number of negative comments has a significant negative impact only for the online sample and when the comments are on the subsequent feedback pages. The non-significance of the revealed impact of the number of negative comments on the first feedback page despite its stated significance (obtained through the self-explicated ratings) can be attributed to the low
number of negative comments compared to positive comments. In the designed DCE, corresponding to most real crowdsourcing marketplace provider profiles and developed through the pre-DCE survey, the number of positive comments (i.e., 3 for low, 15 for high) is much more than the number of negative comments (i.e., 1 for low, 6 for high). Hence, the negative feedback reviews can be simply neglected or “read over” by potential customers, when evaluating provider profiles. H6a and H6b are not supported.

The significant negative impact of the number of negative comments on each subsequent feedback pages (at least in the crowdsourcing marketplace sample) has an important implication. The participants pay more attention to the contents of feedback comments on subsequent feedback pages in the lack of highly visible feedback components on the first, main profile page (e.g., the number of feedback reviews and the average weighted rating). Another interesting finding is that there is no significant effect on customers’ decision-making of the number of positive ratings or the number of positive comments on subsequent feedback pages. Thus, the participants have considered almost exclusively the feedback ratings and comments that appear on the first feedback page. Since the research participants were encouraged to spend enough time for evaluating provider profiles, it is unlikely that the feelings of time pressure have caused them to ignore the rest of the feedback ratings and comments. Along with the other findings discussed above (especially regarding to H3), this suggests that customers consider online feedback components based on the degree of their immediate visibility. Less visible components (e.g., feedback ratings or comments which appear on subsequent feedback pages) require deeper engagement and active monitoring of customers [63] and, thus, they are often ignored. The exemption appears to be negative comments on subsequent feedback pages for which, the participants actively looked and accounted for them in their choices.

We hypothesized that the number of positive textual comments has a larger effect compared to the number of positive numeric ratings (H7). Indeed, the maximum elasticity of the number of positive comments on the first feedback page (0.09 and 0.03 for the lab and online samples respectively) is larger than the maximum elasticity of the number of positive ratings on the
first feedback page (0.08 and 0.02 for the lab and online samples respectively). Furthermore, the average ratings of importance assigned by the participants to these two feedback components (13.08 for the number of positive comments and 10.45 for the number of positive ratings) also support this hypothesis. As above, we attribute this mainly to the strength of text comments to give reasons and tell stories (i.e., telling why the experience was good or bad, not just that it was good or bad). H7 is supported.

With H8a-c, we hypothesized significant synergistic interactions between the number of feedback reviews and the average weighted rating, the number of positive comments, and the number of negative comments on the first feedback page. We expected this mainly because of the argued consistency effect between the corresponding information components [34, 37].

The results show that the effect holds for the first interaction; H8a is supported by the data. The interaction effect postulated in H8b was supported by the crowdsourcing marketplace sample, but not the university lab sample (one might suspect that this could be because the students evaluated much more profiles per person, and might have hence invested less effort). Unsurprisingly in the light of above results (H6a not supported), H8c was not supported by the experiments’ data.

![Figure 4. Importance ranking of feedback information components specified in post-experiment survey](image)
Figure 4 illustrates the rank order of the importance of feedback information components, as specified by the lab participants in a post-experiment survey. The importance rankings are also supportive of our theoretical model, as the majority of the participants have considered average weighted rating, individual comments, individual ratings and other factors respectively important for their decision-making.

The controlled experiment design and the replication of study in two different settings (i.e., controlled laboratory environment and crowdsourcing marketplace) allowed us to increase the external (and internal) validity of our results. While some experimental research have been conducted to evaluate the impacts of online feedback in electronic marketplaces in general [e.g., 64, 65], there is no experimental research investigating the impacts of different feedback information components on customers’ choice decisions. Through these experiments, we could exclude factors outside the scope of our research interest (e.g., past experience of customers with some providers, which is not reflected on the provider profiles). Such irrelevant factors could confound the impacts of the attributes of our interest.

7 Conclusions

This paper presents our theorizing about the impacts on feedback information of provider profiles in crowdsourcing marketplaces, the results of two discrete choice experiments (DCEs) to test, and our interpretation of these results. We conducted two DCEs, one in a controlled laboratory and one online on a crowdsourcing marketplace. We designed the experiments to investigate the main and interaction effects of eight major information components of crowdsourcing marketplace provider profiles on the choice decisions of customers, given the distinct characteristics of these components. The empirical results largely support our set of theoretical hypotheses. The findings of our study have several implications for research and practice of crowdsourcing marketplaces (and electronic marketplaces in general).
7.1 Implications for research

This study contributes to research on, and theoretical understanding of, crowdsourcing marketplaces.

This study is one of the few that have investigated the impact of provider feedback on customers’ choice decisions. Most previous studies have evaluated the impact of some feedback information components on customers’ trust, price premiums, and bid prices [e.g., 27, 30], but not customers’ choices in favor of a provider. We considered the direct as well as interaction effects of eight feedback information components of provider profiles that are common across industry-leading crowdsourcing marketplaces. We found several main and interaction effects of feedback information components to be significant with these not having been previously reported, but important for future theorizing and model development. The findings of this study contribute to our understanding of decision-making in the context of crowdsourcing marketplaces.

Furthermore, we used state-of-the-art developments in experimental research design (i.e., DCEs with Pareto optimal choice sets and minimum aberration fractional factorial design) and analysis (i.e., nested logit). DCEs are highly regarded for their power to realistically replicate (or create) choice situations in an effective manner [18]. This is especially important as some major concerns exist regarding the reliability of findings when “stated preferences” (i.e., survey about speculated behavior) are used instead of actual “observed choices” (i.e., observations of actual behavior) in the analyses [66]. As such, the paper also contributes to the development of improved research methods for topics of interest to IS researchers.

7.2 Implications for practice

This study also has practical implications for practitioners working on or with crowdsourcing marketplaces (i.e., service providers and crowdsourcing marketplace managers).
Especially, our study highlights which feedback information components on crowdsourcing marketplace provider profiles are actually considered by customers when selecting a provider, and what is the relative importance of these components based on their characteristics. This is helpful for both crowdsourcing marketplace managers regarding their design decisions as well as providers in adopting a strategy for managing their profile information to improve their recognition. The model might also be helpful to new customers suggesting to them the profile information that most of customers consider relevant in their selection of providers.

Our results indicate that feedback on the first feedback page has a much larger impact compared to feedback on the subsequent feedback pages of a provider’s profile (even if there is little time difference between their submission to the profile). Globally well-known crowdsourcing marketplaces do not provide any search capabilities for feedback reviews, and their current sort functionalities only result in listing all positive reviews (according to the associated ratings and not the comment contents) at the top, and unrated projects at the end. This virtually implies inaccessibility of many feedback reviews that fall between the two extremes. As such, crowdsourcing marketplace design (e.g., the number of reviews shown per feedback page) has a significant impact on customers’ decision-making, even if the same set of reviews is presented on a provider’s profile. We suggest that crowdsourcing marketplace managers should aim to reduce the inefficiency of customers’ decision processes induced by crowdsourcing marketplace design. One possible way to do so is to integrate content analysis capability and to group feedback reviews by their frequency rather than their recency. For example, some well-known online retailers (e.g., Amazon.com) have recently introduced such content-based aggregation of feedback reviews (e.g., “X reviewers made a similar statement”). Our study implies that such mechanisms are indeed helpful if they are embedded by design, because otherwise “older” information (i.e., on subsequent feedback pages) is not effectively used by customers.

Furthermore, the cost of evaluation of feedback information components has a strong, direct relationship with the likelihood of customers considering these components in their
evaluation. In fact, the cost associated with evaluating a feedback component decreases the importance of the component in customers’ decision-making. Given the potentially high level of similarity of the most visible feedback components (e.g., average weighted rating) among crowdsourcing marketplace providers, the crowdsourcing marketplace managers are encouraged to design strategies which reduce the cost of evaluation of less visible information components. These strategies may include a wide range of services, such as improved sort and categorization, smart aggregation of individual ratings and comments, and hiding irrelevant information components, depending on the requirements of customers.

7.3 Limitations

This research has some limitations. The research was conducted with a limited set of attributes. Specifically, we did not account for other information components, especially those that are generated by providers themselves. Hence, we do not claim to provide a holistic model of all factors that might impact on crowdsourcing marketplace customers’ decisions. As the nature of statistical research, the findings are primarily relevant for “the average case”, we did not discuss or study extreme cases or differences between individuals or groups. The data samples both were biased toward “Western” demographics.

Acknowledgements

This research was supported by the UNSW Australia Business School grant PS29546. The authors are grateful to Associate Professor Aybuke Aurum and Professor Claes Wohlin for their feedback and support. The authors would also like to express their appreciation to Professor Fred Mannering and Professor Simon Washington, for their valuable advice on the application of discrete choice models.

References


