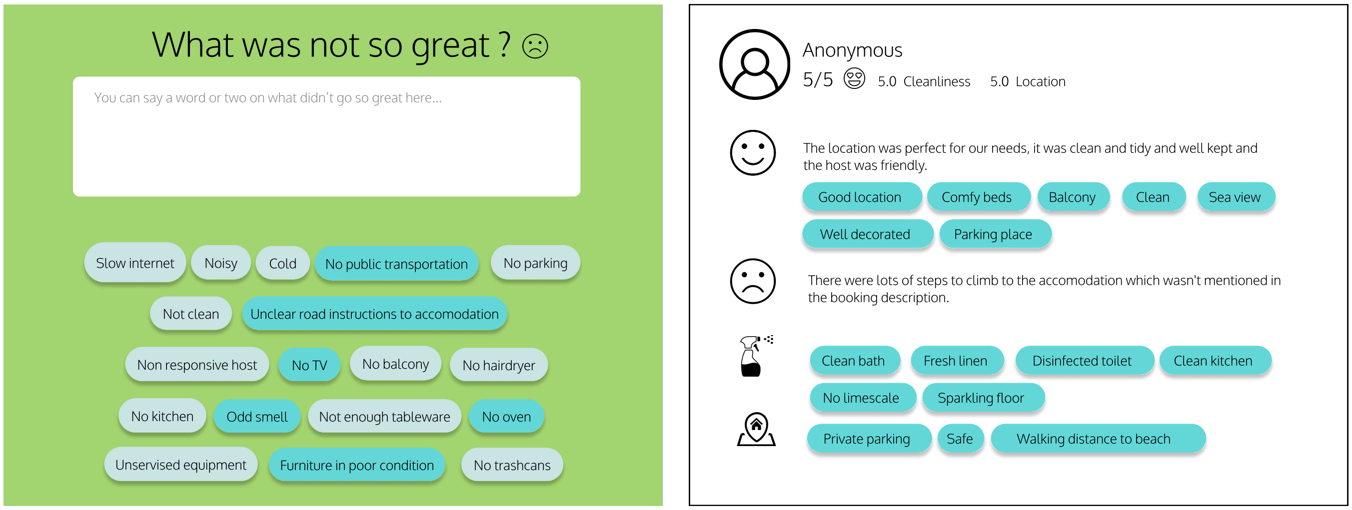
Less Typing, More Tagging: Investigating Tag-based Interfaces in Online Accommodation Review Creation and Perception

Teodora Mitrevska
LMU Munich
Munich, Germany
teodora@mitrevska.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

Christina Schneegass
TU Delft
Delft, Netherlands
c.schneegass@tudelft.nl



(a) Tag support for review creation. **(b) Tag-based interface for review perception.**

Figure 1: We present a *tag-based* prototype to support the generation and reception of accommodation reviews: During (a) review creation, predefined tags complement the free text entry field to facilitate memory recall and ease review input. When (b) viewing the final review, tags help users quickly perceive important information.

Abstract

When booking accommodations such as hotel rooms or vacation houses online, users heavily rely on the experiences of prior customers through reviews. However, such reviews are often short or lacking important details since most UIs limit users to text-only input, making review creation laborious and time-consuming. The quality of the review can therefore vary, depending on the skill and intrinsic motivation of the reviewer. Instead of relying on these two variables, we explore the effects of a tag enriched UI, both for creating and presenting review information. In the process, we evaluate *tags* – short descriptive text snippets such as "centrally located" – as additional input components to open text fields and rating scales. These tags aim to trigger users' memory of details and experiences and support them in creating comprehensive reviews without having to generate long texts. In an initial user study, we asked participants to generate reviews with and without the tag support. In a second user study, they evaluated reviews created with and without tags. Our results show that tags were perceived as helpful in creating reviews and increased the level of detail in the experiences reported. We discuss the implications of review quality, helpfulness, and potential limitations.

CCS Concepts

• **Human-centered computing** → *Human computer interaction (HCI)*; **Empirical studies in HCI**.

Keywords

online reviews, user interface design, user experience

NordiCHI 2024, October 13–16, 2024, Uppsala, Sweden

ACM ISBN 979-8-4007-0966-1/24/10
https://doi.org/10.1145/3679318.3685369

ACM Reference Format:
Teodora Mitrevska, Sven Mayer, and Christina Schneegass. 2024. Less Typing, More Tagging: Investigating Tag-based Interfaces in Online Accommodation Review Creation and Perception. In *Nordic Conference on Human-Computer Interaction (NordiCHI 2024), October 13–16, 2024, Uppsala, Sweden.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3679318.3685369

1 Introduction

Online accommodation booking systems such as *Airbnb*[1] or *Booking.com*[2] have changed the way people travel. Such online platforms are easy to access, offer plenty of options, and have a simple booking process. While the presentation of service is important, user-generated reviews have become a crucial factor in the success of accommodation providers [24, 36, 41]. They contribute valuable insights into an accommodation beside the information made available by the platform and help future customers in their decision-making process [15, 23]. While customers value access to detailed experience reports, writing reviews demands an investment of time and effort to report on one's own experience. Therefore, accommodation reviews often lack sufficient detail and breadth, which are essential for creating useful and high-quality reviews for the next customer [8, 14, 19]. Booking platforms do their best to create persuasive user interfaces (UIs) that support users in generating reviews. Yet, many review forms still deploy rating scales, which are easy to answer but lack details to justify the rating. What explains a rating of 6 out of 10 points for the category "Facilities"? It can be anything from a disappointingly small swimming pool to a wheelchair-inaccessible hotel lobby or an ungrateful guest who did not like how the toilet paper was folded. A similar effect is caused by text fields as an input source. In fact, text fields are often provided as input for a general review i.e answering the question "How was your stay?". While the question itself is very vaguely phrased, the UI is not supporting the user with any guidance for the answer. Therefore, the quality of the reviews can vary greatly, as writing extensive textual reviews and identifying personally relevant information can be tedious and time-consuming. Furthermore, as mobile devices are nowadays predominant, supporting users in entering their reviews with low effort and quickly is essential.

Tag-based interfaces have been studied in many application contexts (cf. [7, 11, 34] and can deliver a wide range of options, allowing it to be a fast and detailed input method. Here, the interface provides a set of predefined options for the user to pick from, see Figure 1. These interfaces have been to benefit over traditional selection and input methods. For instance, [34] showed that tag-based interfaces enhance searching for specific objects, and Ravendran et al. [25] showed the benefit of tag-based interfaces for mobile banking. On a small scale, Airbnb deploys tags to provide more details on specific ratings, such as why the user selected "moderately comfortable" for the question "Was the place comfortable?" Today, these tags are only used for specific categories, are limited in number, and are not visible to the next user on Airbnb. Thus, many open questions remain regarding their helpfulness for review creation and later during review perception. We see several important benefits in *tag-based interfaces*: a) They provide users with an **easy** and **fast** way of reporting on their experiences **with minimal typing** required, by supporting the **recognition rather than recall** principle [1] and therefore positively influencing **user experience** (UX) [26, 27]. b) They function as a **memory aid**, triggering the recall of experiences, hereby creating **more detailed** and, thus, **more helpful** reviews. c) They create a **concise overview** of how prior customers reasoned their reviews' ratings, by controling the relevance of the content. d) They aid the structure of information, helping the user process the content faster [21].

This work explores using tag-based interfaces to improve the review interfaces from a review creation and reception perspective. We implement a prototype that deploys tags complementary to rating scales and free-text entry fields (we call "*tag-UI*"). In this case, tags are pre-defined short text snippets that express certain frequent experiences with an accommodation (such as "no balcony" or "centrally located") as an optional addition to each free text entry field and rating scale.

We evaluated our prototype in a between-subject user study (N = 59). Here, we explore how the tags are perceived as supplementary UI when creating reviews compared to only the traditional free-text answer fields and rating scales. While easy and fast creation of reviews is important, they also need to deliver the information to the reader. Thus, in a second within-subject study (N = 59), we evaluated the quality of the reviews created in the first study. We compare how users perceive the reviews from the two conditions, reviews with and without supplementary tags, concerning review quality, quantity, and overall helpfulness.

Our results show that tags ease the review creation process, and their availability is perceived as helpful compared to traditional review formats. The second study confirms that for review perception, tags provide several benefits when provided as an addition to text or instead of text. Reviews containing tags are perceived as containing more diverse information while being just as trustworthy and detailed as text-only reviews. We discuss the implications of tags on the user experience of creating and perceiving online accommodation reviews and potential limitations. By providing valuable insights into how this UI component influences the experience, we hope to contribute to the growing body of research and support developers and designers in creating review forms that encourage users to write detailed and useful reviews.

2 Related Work

2.1 What constitutes a "helpful" review?

It can be debated if, for online accommodation reviews, the review quantity or the review quality holds more significance. While some argue that a higher number of reviews signifies popularity and credibility, others emphasize that the quality of reviews holds greater importance in determining a property's overall value [24]. Research indicates that while the number of reviews impacts booking rates, it is the quality of these reviews that significantly influences guest satisfaction and the likelihood of return bookings [4, 40]. Sparks and Browning's study [30] further highlights that specific and detailed feedback from reviewers rather than simple ratings or vague comments enhances the credibility and trustworthiness of online reviews, leading to increased booking intentions.

To conduct a comprehensive evaluation of user reviews, it is essential to establish criteria for defining quality information. Churchill [10] proposed measures for assessing information quality, including information depth, breadth, factuality, relevance, and credibility. Building on this concept with a focus on online consumer reviews, Filieri [13] defined the following four variables:

Information depth and breadth: This refers to the extent to which a review provides detailed coverage of various aspects

[1]Airbnb - https://www.airbnb.com/, last accessed 2024–11–02
[2]Booking.com - https://Booking.com/, last accessed 2024–11–02

related to a product or service. In the context of accommodations, such a review may discuss room cleanliness, breakfast quality, and other details.

Information factuality: It is based on presenting specific facts about a product or service.

Information relevance: This criterion assesses how applicable and helpful a review is for a particular task or situation, considering various consumer needs [38].

Information credibility or accuracy: It gauges the extent to which users perceive a recommendation or review as trustworthy and true [9, 35].

Filieri [13] argues that reviews with greater detail, completeness, accuracy, factual basis, and relevance to consumer needs are perceived as more helpful by potential buyers in evaluating product or service quality. However, it is worth noting that distinguishing between truly positive and less-than-positive reviews may rely on subtle differences in wording, as observed by Bridges and Vásquez [5]. Lastly, researchers have found that review creators could generate more helpful reviews by highlighting important review attributes [39]. In conclusion, review quality and quantity are both important factors for prospective guests when selecting rental properties. Nevertheless, detailed and relatable feedback is particularly valuable in providing insights into a property's quality and suitability for individual needs.

2.2 Tags in User Interfaces

Tags as user interface components are not novel. They have been applied in numerous domains, including mobile banking [25, 27], video [7] and image annotation [16], lifelog retrieval interfaces [31] and other application contexts. Ravendran et al. [25] showed that tags in a novel mobile banking application enabled more customization and enhanced the user experience compared to the traditional system. Further underscoring the significance of tag-based interfaces, Trattner et al. [34] examined the efficacy of social tags in accessing information. They specifically looked into tag clouds as a medium for information visualization and access. Their user-centric approach revealed that tag-based browsing interfaces were far more efficient and satisfying for users than traditional search methods. Additionally, research has shown that tags turned out to be helpful not only for giving recommendations and improving search in social tagging systems but also for enhancing information access by navigating [11]. Further on, Ravendran et al. [27] showed that tag-based interfaces as selection mechanisms improve usability and user satisfaction in a banking app. Regarding the multi-functuality of tags, Vig et al. [37] showed that tag-based interfaces for searching images outperform their traditional text-based counterparts, which can help add another dimension to the consumption of the user-created content. Last but not least, Banda and Bharadwaj [3] showed that not only user have benefits but also systems; in their case, recommender systems have better performance when incorporating collaboratively generated tags. Moreover, Kailer et al. [17] showed how to improve tag quality by rating tags itself. Such rated tags can be directly incorporated into the user interfaces themselves, presenting promising results in increased decision quality and a decreased decision making effort [18].

3 Designing Tag-based UI Prototype

We developed a tag-UI prototype and a control prototype to investigate the impact of tags on the review writing and review reading quality. In the tag-UI prototype (see Figure 2a and 2b), tags are provided complimentary to rating scales and free-text entry forms to support the user in the review creation process.

3.1 UI Analysis of Two Booking Platforms

To get a better understanding of the state-of-the-art commercial accommodation booking systems and their review forms, the following section will present a brief analysis of two major platforms, *Booking.com* and *Airbnb*. The analysis focuses on UI elements and content guiding the review creation process[3].

The review system of both platforms starts with user profiling, creating a persona based on trip details to enhance relatability and decision-making confidence. Booking.com asks about the purpose of travel (business or leisure) and the size of the party, while Airbnb skips this step and does not gather additional traveler information. Both platforms use rating scales to gather overall assessments of the accommodations. Booking.com employs a numerical scale from 1-10, while Airbnb uses a 5-point star rating scale. Besides the rating of the overall stay, Booking.com also allows travelers to rate various aspects of the accommodation using emoji-based scales (sad, neutral, happy, very happy). These aspects include but are not limited to Staff, Cleanliness, Comfort, and Location. Airbnb follows a similar approach, using star scales for individual aspects (e.g., Check-in, Cleanliness, Accuracy). In both platforms, additional information on the reviewer's accommodation assessment is gathered through open text input fields. Airbnb keeps it simple, asking for a general public review and a private note to the host. In contrast, Booking.com separates open feedback into what the traveler liked and disliked, prompting reviewers to mention both positive and negative aspects. Additionally, both platforms include platform-specific elements for assessing additional accommodation attributes. Booking.com introduces "Bonus questions" about the surroundings, amenities, and special requests fulfilled. The more questions reviewers answer, the more pop up, making it a voluntary content contribution. Similarly, Airbnb presents complex scales with expanded choices for feedback. For example, when rating space equipment, selecting a value leads to a list of specific items to assess, allowing reviewers to provide detailed input.

As already mentioned in the introduction, Airbnb already deploys tags as reasoning options for specific rating questions, such as "Was the place comfortable?" or "Was the description accurate?". Here, a list of around six to ten tags is shown, adjusted to the respective rating scale item the user clicks. For example, for the question on the comfort of the accommodation, the tag "Good temperate" is displayed when the user selects that the place is extremely or very comfortable. If the user selects moderately comfortable, a tag saying "temperature too hot or too cold" is displayed. In the case of Airbnb, the tags are only used for rating scales after the written review has already been typed. The tags are not displayed to future users, thus providing very little value to the customers.

[3]Due to the possibility of changes in the interface by the time this publication is reviewed, screenshots of the process from the analysis day can be found in the supplementary material folder.

(a) Ratings + tags - location **(b) Ratings + tags - cleanliness** **(c) Control condition**

Figure 2: Screenshots from our *tag-UI* prototype showing the complementary tags to the rating scale for the topics (a) location and (b) cleanliness compared to (c) the control condition.

In conclusion, both platforms utilize many stand-alone rating scales that do not offer sufficient detail on how the rating was derived. Undirected open comment fields further provide no guidance on what the user is supposed to write, which can lead to comments that do not justify the overall rating or even conflict with each other. Tags, when applied, are limited in number, only used for very specific questions, and are not meant to be displayed to the next customer.

3.2 Design

The analysis of existing platforms revealed that the review process typically consists of three steps that reviewers are required to complete, along which we built both our prototypes[4]:

Step 1: Providing an overall rating score for the accommodation: Our no-tag control condition utilizes a 5-point star rating scale as used by Airbnb to assess customer's overall impression of their stay. In contrast, the tag-UI prototype deploys a 5-point emoji rating scale as research has shown that emoji scales can convey emotions better [33], reduce bias [20, 32], increase engagement [2], and are more intuitive [28]. No tags are presented at that stage yet.

Step 2: Leaving a textual review: Our control prototype presents one free text entry field asking users to provide a public review. In our tag-UI prototype, we implement two text fields on separate screens, one for positive and one for negative comments, supplemented with tags (see Figure 1a). The tags are implemented as buttons that change colour upon selection. In this step, the tag-UI prototype differentiates from Airbnb's interface by combining tags and text fields as an input source. This way, we assume that the users have a chance to structure their feedback by reviewing the negative and positive sides of their stay separately, but also increase the detail level of the produced content by choosing tags.

Step 3: Rating additional amenities/facilities: Similar to Step 1, our control prototype uses 5-point star ratings positioned together on one page while the tag-UI continues the emoji rating scale. Upon selecting a rating, supplementary tags that fit the respective rating appear below the scale (see Figure 2a and 2b). Similar to the Airbnb concept, the users can select as many tags as they want.

Our control condition replicates the status quo review process (comparable to Airbnb and Booking.com; see analysis presented in Section 3.1). By comparing our tag-UI to our own control prototype instead of existing booking platforms, we aim to minimize effects caused by brand and experience bias and ensure that our effects originate in the tag-UI as the main difference. By comparing these two systems, we aimed to assess the effectiveness and user-friendliness of our tag-UI for online accommodation booking.

Since this study is an initial exploration of the tag-UI concept for accommodation reviews, we limited the number of tags and interactivity. The tags used in the study were derived from common phrases used in Airbnb and Booking.com reviews and kept static across the two studies. We limit the phrases within one tag to a maximum of five words to keep them brief and easy to parse and present a maximum of 20 tags simultaneously. We acknowledge that this interaction has limitations and should be further developed in future work to offer more adaptivity and personalization. We present our considerations in the discussion section.

We implemented both prototypes as web applications using React (version 18.2.0). For the user studies, the interaction with the prototypes is logged as a JSON object and stored in a Firestore database on Google Firebase[5], connected to the frontend via NodeJS. The prototypes were hosted on Netlify[6].

4 User Study 1 - Review Creation

To compare our tag-UI prototype to the status-quo prototype, we conducted an online survey using Qualtrics with Prolific[7] integration. In a between-subject study, we aim to investigate whether our experimental condition, the tag-UI, can deliver a better UX during the review creation process compared to the control condition. We will specifically analyse users' perceptions of the tags and their satisfaction with the reviews they created. We further assess how reviews from both conditions differ regarding their length and level

[4]Screenshots of the prototypes can be found in the supplementary materials.

[5]Google Firebase – https://firebase.google.com/, last accessed 2024–11–02
[6]Netlify – https://www.netlify.com last accessed 2024–11–02
[7]Prolific – https://www.prolific.co/, last accessed 2024–11–02

of detail. The subjective helpfulness of the reviews will be evaluated in the second user study. Thus, we postulate the following main research question for Study 1:

RQ1: How do users perceive the tag-UI prototype during the review creation process compared to the control prototype?

4.1 Procedure

At the beginning of the study, participants were welcomed and informed about the study's process, goal, duration, and data management. We assigned each participant a randomized ID for anonymous data processing and complied with the General Data Protection Regulations (GDPR). Every participant provided informed consent to start the study. As the first step, we briefed participants to think of their recent travel experience and asked them to interact with either the experimental or control prototype, which we randomly assigned for our between-subject test. In both conditions, participants were asked to imagine they are creating a review for their latest travel accommodation to be shown to other customers interested in this accommodation. After the prototype interaction, they were redirected to a survey. We ask them to provide information on the recent trip they had in mind while creating the review to monitor potential confounding variables that could affect review length and quality (e.g., the length of the stay could influence the number of details they recall). In addition, we asked participants to describe their review creation habits, such as the frequency of writing reviews, the time they usually invest, the content they report, and their use of reviews for their booking.

In the main survey, we asked participants to report their experience with the prototypes with complementary quantitative and qualitative measures. For every step of the review process, they indicated their impression and understanding of the UI in their own words. To summarize their experience, they finally indicated in four Likert ratings (5-point, from 1 = "strongly disagree" to 5 = "strongly agree"), (a) how well they were able to capture their experience with the accommodation using the UI, (b) how satisfied they were with the level of detail of their review, (c) how satisfied they were with the length of their review, (d) and if they believe their review will help the next user make an informed decision on the accommodation.

For the tag-UI prototype, we further included survey questions on users' perceptions of the tags. Since the full versions of established UX or UI questionnaires were inadequate for this use case, we created more suitable statements that serve our case, as well as adjusted some from well known UX questionnaires [6, 29]. We asked participants to rate the following statements on a 5-point Likert scale (1 = "strongly disagree" to 5 = "strongly agree"):

(1) "Tags are easy to use"
(2) "Tags help to express my opinion better"
(3) "Tags are fun to use"
(4) "There were sufficient amount of tags"
(5) "Tags are time consuming"
(6) "I would like to use tags frequently for reviews"
(7) "Tags help me finish review faster"
(8) "Tags are distracting"
(9) "Tags help me remember details about my stay"

To rule out UI misconceptions, we also asked participants to describe their understanding of the tags and their function in their own words and indicate if they considered the tags optional or mandatory. Further, they were asked to report any questions or concerns with this interface.

Finally, we asked our participants for basic demographic information, including gender, age, educational degree, current occupation, number of trips per year, and their experience with popular accommodation booking platforms. A final open comment field for remarks concludes our first study.

4.2 Participants

We recruited participants via Prolific from the UK, US, and Canada to ensure sufficient English language proficiency to participate in our study. As a requirement for participation, we expect people to have booked travel accommodation and travelled in the last three months to be able to relate to recent experiences. The study ran in January 2023 and lasted around fifteen minutes, and participants were compensated according to Prolific recommendations with 10 USD/h.

We recruited 60 participants, of which one had to be excluded due to incomplete data, and aimed for a diverse sample balanced in gender (32 male, 27 female) and age (range 18-74, M = 36.2, SD = 14.55). An apriori power analysis was conducted using G*Power version 3.1.9.6 [12] to determine the minimum sample size. Results indicate the required sample size to achieve 80% power for detecting a large effect at the significance criterion of $\alpha = .05$ was N=54 ($N = 27$ per group) for a Mann-Whitney U Test. For a Kruskal Wallis test, again with an expected 80% power and large effect, our sample size calculation results in a minimum $N = 52$. Thus, our sample size of 59 is sufficient.

4.3 Results

We checked all data for normal distribution and equality of homogeneity of variance and reverted to the respective alternative test if necessary.

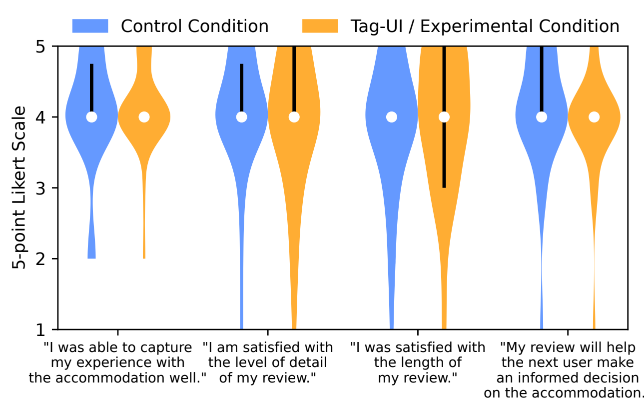


Figure 3: The four UX questions for the tag-UI and control conditions were rated on a scale from 1 (strongly disagree) to 5 (strongly agree). The white dot indicates the median and the thick black line indicates the interquartile range.

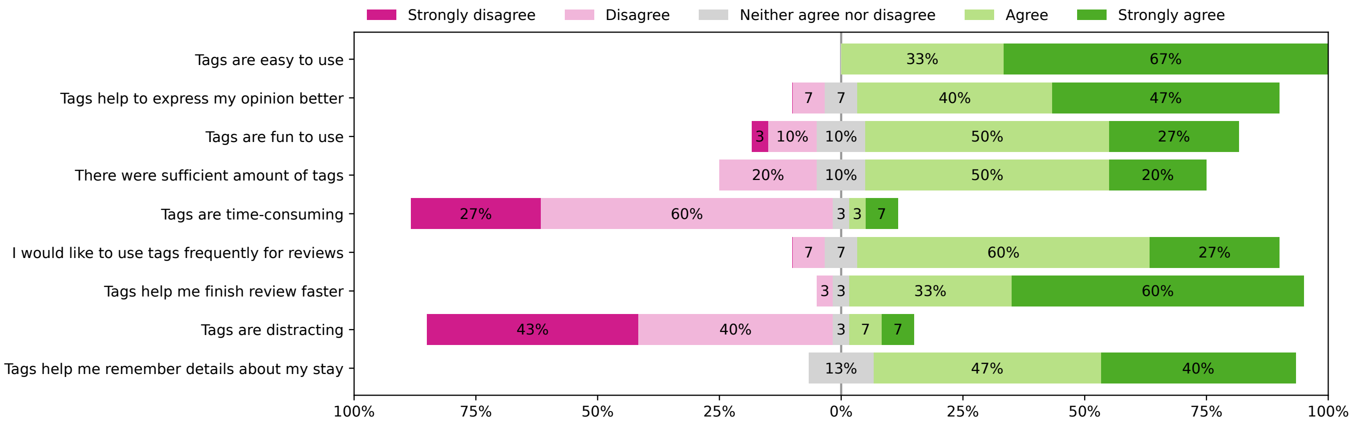


Figure 4: Overview of Likert ratings of the users' experience with the tags in our tag-UI.

4.3.1 Review Habits. The majority (72%) of participants indicate providing reviews rarely or sometimes on a scale from 0 = "never" (over rarely, sometimes, often) to 5 = "always"), and only four indicate always leaving a review. Between our two groups, there was no significant difference in their frequency of providing reviews (Mann-Whitney U test, $U = 805.5, p = .302$). Regarding the time they invest in writing these reviews, around half of the participants (31/59) stated to take 1-5 minutes, and only some participants indicated to spend 6-10 minutes (16/59) or less than 1 minute (10/59). Yet, when asked how strongly they usually rely on reviews when selecting an accommodation, our participants rated their importance very high ($Md = 4$, 5-point Likert scale from "not at all" to "very strongly").

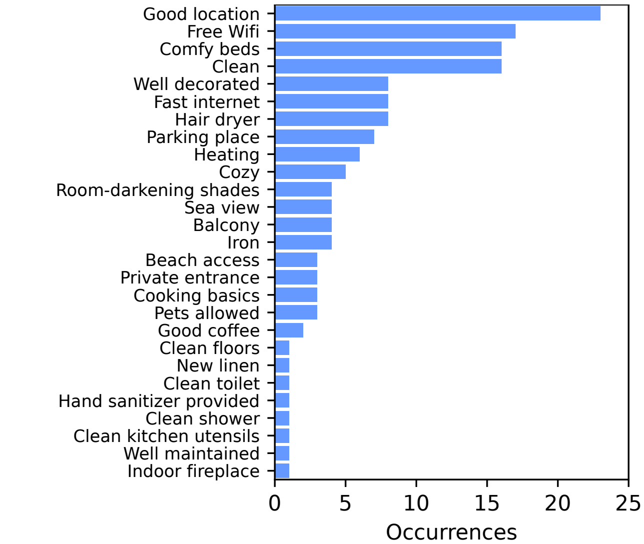
4.3.2 User Experience Overall. We compared users' ratings between the experimental and control conditions, specifically on four UX items. No significant difference in the four questions on review creation experience (Mann-Whitney U Test) concerning participant's ability to capture their experience with the respective UI ($U = 423$, $p = .825$), how satisfied they were with the level of detail of their review ($U = 405.5$, $p = .621$) and with the length of their review ($U = 407.5$, $p = .656$), and if they believe their review was helpful ($U = 359.5$, $p = .172$). Figure 3 shows that the ratings for both prototypes were very high, with a median of 4 across all questions and groups.

4.3.3 Perception of the Tag-UI. In our experimental condition, most participants considered the tags an optional feature (23), while some considered them mandatory (6). When the tags were presented in addition to free text, participants perceived the tags as, among others, "enrichment to the review", "highlights of the trip", or "buzzwords". One participant described their understanding of a tag, saying, "the tag allows for quick reviews, while the text boxes allow for a more in-depth review" (S1P23). In addition to rating scales, participants described that they "understood the tags as additional helpful indicators to explain further my choice for the rating scale; the rating scale is guided by the examples offered in the tags, not vice versa." (S1P18). Most participants stated similar perceptions, describing that the tags are very helpful "[...] to explain the reasoning behind your rating" (S1P25). Overall, all descriptions of the participants fit our intention, showing that participants understood the usage of the tag-UI.

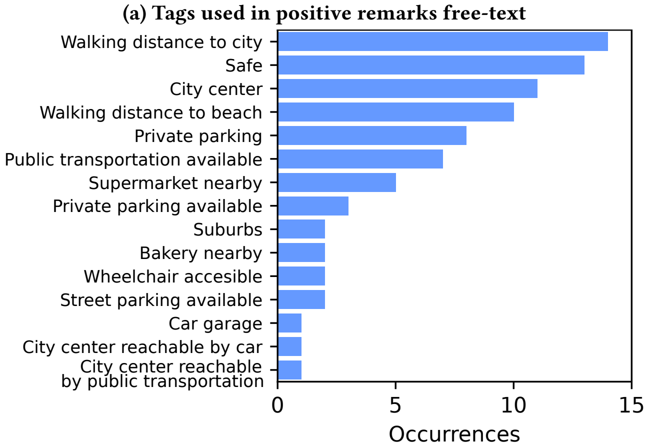
We further asked our participants to rate the tag-UI on several statements describing their usability, UX, and helpfulness in creating reviews (5-point Likert scale from strongly disagree (=1) to strongly agree (=5)). The results of these ratings can be seen in Figure 4. In summary, the tags are considered easy to use ($M = 4.67$, $Md = 5$, $SD = 0.48$), helpful in expressing one's opinion ($M = 4.27$, $Md = 4$, $SD = 0.87$), neither time-consuming ($M = 2.03$, $Md = 2$, $SD = 1.03$) nor distracting ($M = 1.93$, $Md = 2$, $SD = 1.17$), helpful to create a review faster ($M = 4.5$, $Md = 5$, $SD = 0.73$), and supporting the recall of details about the trip ($M = 4.27$, $Md = 4$, $SD = 0.69$).

4.3.4 Tag Quantity & Review Length. We analyzed how many tags were used in addition to the free text entry fields describing the likes and dislikes. In the field where users were asked to report positive remarks on the accommodation, more tags were used ($M = 5.1$, $SD = 2.86$, range 0-11) compared to the section asking for negative remarks ($M = 1.77$, $SD = 2.14$, range 0-11). The most commonly selected tags in the positive remarks section were "Good location", "Free Wifi" and "Comfy beds" (see Figure 5a), whereas "No parking", "Noisy" and "No kitchen" were most selected in the negative remarks section. The tags that were never selected from the pool for the first positive remarks section were "Playground" and for the negative section "No trash cans". In the location section, the most frequently selected tags were "Walking distance to city", "Safe" and "City center" (see Figure 5b), whereas "Disinfected toilet", "Clean bath" and "Fresh linen" were the most frequently selected ones in the Cleanliness section.

When analysing the additionally provided textual input, a Kruskal-Wallis test revealed a significant difference in the review's word count between our status quo interface and our tag-UI ($H(1) = 8.001$, $p = .005$). In the control condition, participants phrased their reviews with significantly more words ($M_{Ctl} = 35.31$, $SD_{Ctl} = 31.84$, Range 6-114) compared to what they used in the tag-UI condition for the like and dislike text field combined ($M_{Exp} = 10.65$, $SD_{Exp} = 21.34$, Range 0-99). There was no notable difference between the length of the textual input in the positive and negative remark field ($M_{pos} = 8.6$, $SD_{pos} = 11.86$, Range 0-56; negative:

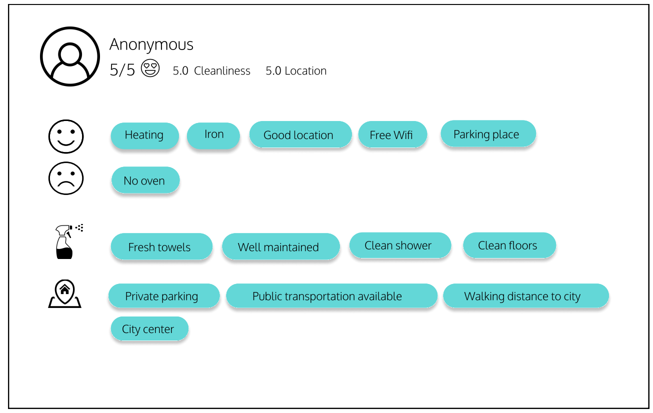


(a) Tags used in positive remarks free-text



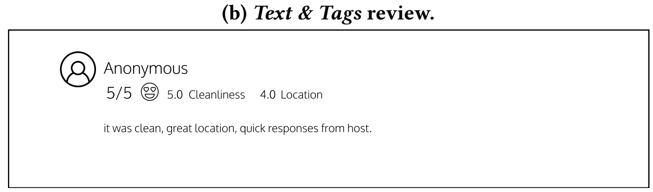
(b) Tags used in location rating

Figure 5: The most frequently selected tags in the tag-UI prototype for the positive free-text entry field (a) and the rating scale for the location attribute (b).



(a) ***Tags-Only*** **review.**

(b) ***Text & Tags*** **review.**



(c) ***Text-Only*** **review.**

Figure 6: The three user interface categories we compared in our second study: (a) showing only tags, (b) a combination of tags with text, and (c) only free text reviews.

M_{neg} = 7.8, SD_{neg} = 10.57, Range 0-43). However, what was interesting to observe was that in our experimental condition that showed the tag options, nine out of our 30 participants did not enter written text at all, only tags. We discuss these results together with the results from the second study in Section 6 – Discussion.

5 User Study 2 - Review Evaluation

We conducted a second user study to understand if the perception of reviews with tags is better than that of traditional reviews without tags. We aim to support the initial expectations that our tag-UI prototype produces more helpful reviews than the traditional, no-tags control prototype. In the survey, we use the reviews produced in the first user study to evaluate the users' subjective perception of the reviews. From our two factors, tags and text, emerge three possible display options: *Tags-Only*, *Text-Only*, and *Text & Tags*. We evaluate the three conditions using a randomized within-subject design to allow participants to compare them. We postulate the following research question:

RQ2: How do users perceive review containing *Tags-Only*, *Text & Tags* and *Text-Only*?

Specifically, we will analyze the users' assessment regarding the reviews' general understandability and information quality, quantity, and relevance.

5.1 Apparatus

We embedded the reviews produced in the first study in a standardized interface, see Figure 6, and presented them with anonymized user information, i.e., without a name and with a profile picture placeholder, to avoid any biases caused by the review authors. From the first study's no-tag condition, we collected 29 Text-Only reviews. In the tag-based interface condition, we collected 21 reviews with text and tags and nine where users selected the tags but did not write complementary free text. Thus, we split this group into a *Text & Tags* and a *Tags-Only* condition. We randomized the reviews generated within each of the three conditions so that every participant would see one review from each condition.

Tags-Only review. This category was generated by participants who chose not to type any text, but rather use tags and rating scales to create a review. It displays an overall rating score, cleanliness, and location score. Additionally, it displays the chosen tags in the likes and dislikes section and the tags describing the cleanliness and location aspects of the rental accommodation (see Figure 6a).

Text & Tags review. This category was also produced via our tag-UI prototype by the participants writing a textual review and choosing some tags to go along with it to provide more detail. Identical to the previous category, this one also characterizes the reviews with an overall rating of the accommodation and additional tags and scores chosen in the cleanliness and location sections (see Figure 6b).

Text-Only review. This category was produced by our no-tag control condition since the input fields only allowed the reviewers to enter textual input and choose a rating score on a rating scale. Using this system, the reviewers produced reviews that contain an overall rating score, rating scores for cleanliness and location, and a general textual review of the accommodation rental (see Figure 6c).

5.2 Procedure

At the beginning of the study, participants were welcomed and informed about the study's process, goal, duration, and data management, as well as compliance with the General Data Protection Regulations (GDPR). As a first step, the participants were asked to provide basic information about their trip, such as destination, type of trip (i.e., business trip, vacation, or other), and the traveling party (i.e., colleagues, family, couple). We proceeded by asking questions about the participants' review habits, such as the frequency of writing reviews for accommodations they have stayed in, time spent writing a review and content provided. Further, we were also interested to know how strongly the participants rely on reviews when booking and on which element of the information presented online they rely the most (e.g., pictures of the accommodation, user reviews, textual descriptions, or other).

In the next part of the study, the participants were asked to evaluate the three types of reviews produced by the participants in the first user study. In a randomized conditions order, each participant received one randomly chosen representative of the conditions to evaluate. After every condition, we asked the participants questions on the general understanding of the different pieces of information in the review, perceived trustworthiness, reliability, and helpfulness of the reviews. We also asked about the level of satisfaction regarding each review's information quality, quantity, and relevance [10, 13]. We phrased the questions as statements, allowing the participants to indicate their opinion in Likert ratings (5-point, from 1 = "strongly disagree" to 5 = "strongly agree").

Ultimately, we asked the participants to reflect on all three review samples and provide some qualitative feedback on which review they found most helpful. Finally, we asked participants for basic demographic information.

5.3 Participants

For the purpose of the second user study, we recruited participants using Prolific from the UK, the US, and Canada to ensure sufficient English language proficiency. The participants were chosen based on fulfilling the requirement of having traveled abroad in the past six months, under the assumption that they booked a rental accommodation and, for that purpose, browsed through some consumer reviews. We recruited 59 participants; 30 identified as male and 29 as female. Their age ranged from 23 to 75 (M = 41.1, SD = 13.8). 36% of the participants expressed that they provide reviews *often*. 36% *sometimes*, 19% of them *rarely*, 7% *always*, and 3% *never* write reviews. Most participants (90%) answered that they usually provide an overall rating score of the accommodation. More than half expressed that they *strongly* (54%) or even *very strongly* (14%) rely on user reviews when booking rental accommodation.

5.4 Results

5.4.1 *General Understanding.*

Initial Shapiro–Wilk tests confirmed that the data for *Understandable*, *Helpful in Decision-Making Process*, *Trustworthy* and *Reliable* is not normally distributed; see Table 1. We conducted Friedman tests for all items. For *Understandable*, no significant difference was found (p = .860).

However, for *Helpful in Decision-Making Process*, a Friedman test indicates a significant difference, which the post-hoc test revealed to be between the *Text & Tags* and *Text-Only* condition ($Z = -2.926$, p = .003). **Precisely, *Text & Tags* (M = 4.29, SD = .85) are perceived as significantly more helpful in the booking decision-making process compared to *Text-Only* (M = 3.85, SD = 1.03).**

While a Friedman test indicated a general difference between the conditions for *Trustworthy* (see Table 1), post-hoc tests did not reveal any significant difference ($p > .1$). Thus, our findings showed that **there were no significant differences found in perceived trustworthiness among the three conditions**, see Figure 7a. In fact, the data shows that the information in the samples from all three conditions was found to be rather trustworthy and reliable. There was a big similarity in the distribution of answers noted between *Tags-Only* (M = 3.6, SD = 1.0) and *Text-Only* (M = 3.5, SD = 0.9) for the perceived trustworthiness.

For the item ***Reliable***, a Friedman test showed significant differences; see Table 1. Afterward, we performed Wilcoxon signed-rank tests with Holm–Bonferroni correction applied; however, they showed no significant differences ($p > .05$). Identical results were found regarding the perceived reliability for *Tags-Only* (M = 3.6, SD = 1.0) and *Text-Only* (M = 3.5, SD = 0.9). Compared to the previously mentioned review conditions, the results for *Text & Tags* stood out. For *Text & Tags*, most participants stated that they either agree or strongly agree in finding the presented sample reliable (M = 3.8, SD = 0.8) and trustworthy (M = 3.8, SD = 0.9).

Table 1: Friedman test results comparing participants' ratings of *Text-Only*, *Tags-Only*, and *Text & Tags* reviews. For the effect size, we follow the Cohen's d interpretation that a small effect is .2, a medium effect is .5, and a large effect is .8. * We tested the normality using the Shapiro-Wilk test.

	Friedman test			Effect Size	Normality*	
Item	**df**	χ^2	**p**	**Kendall's W**	**W**	**p**
General Understanding						
Understandable	2	.30	.860	.003	.736	<.001
Helpful in Decision-Making Process	2	10.01	.007	.085	.809	<.001
Trustworthy	2	8.13	.017	.069	.882	<.001
Reliable	2	7.77	.021	.068	.888	<.001
Information Quality & Quantity						
Quality was Satisfactory	2	9.17	.010	.078	.859	<.001
Quantity was Satisfactory	2	19.99	<.001	.172	.887	<.001
Sufficient Detail	2	11.64	.003	.099	.887	<.001
Sufficient Breadth (spanning different subjects)	2	24.86	<.001	.211	.882	<.001
Information Relevance						
Helpful to Evaluate Accommodation	2	12.22	.002	.104	.830	<.001
Helpful to Familiarize with Accommodation	2	8.23	.016	.070	.887	<.001
Easing Decision-Making	2	8.11	.017	.069	.879	<.001
Providing Additional Information	2	14.92	<.001	.126	.892	<.001
Reflecting Subjective Experience	2	2.51	.285	.022	.853	<.001

5.4.2 Information Quality & Quantity. First, a Shapiro–Wilk test confirmed that the data for *Quality was Satisfactory*, *Quantity was Satisfactory*, *Sufficient Breadth* and *Sufficient Detail* is not normally distributed; see Table 1. Thus, we performed Friedman tests, which showed a significant difference between the three conditions in all four items, see Figure 7c. Again, we performed post-hoc Wilcoxon signed-rank tests with Holm–Bonferroni correction applied.

For *Quality was Satisfactory*, the post-hoc test showed a significant difference between the *Text & Tags* and *Text-Only* condition ($Z = -3.385$, $p < .001$). Here, ***Text & Tags* ($M = 3.95$, $SD = .955$) received a significantly higher rating in satisfaction with the quality of the reviews compared to *Text-Only* ($M = 3.36$, $SD = 1.20$).**

In the item *Quantity was Satisfactory*, we found similar effects as in the review quality. The post-hoc test showed again a significant difference ($Z = -3.649$, $p < .001$) between *Text & Tags* and *Text-Only*. Overall, **participants rated the quantity of information in the *Text & Tags* ($M = 3.71$, $SD = 1.08$) higher than in the *Text-Only* condition ($M = 2.98$, $SD = 1.13$).**

For *Sufficient Detail*, the results revealed that there is a statistically significant difference between *Text-Only* with both *Tags-Only* ($p = .019$) and *Text & Tags* ($p = .003$). However, not between *Tags-Only* and *Text & Tags* ($p = .486$). The results show that participants tend to strongly agree that **reviews from *Tags-Only* ($M = 3.31$, $SD = 1.25$) and *Text & Tags* ($M = 3.42$, $SD = 1.16$) consist of a sufficient level of detail**, while they give lower ratings in the *Text-Only* condition ($M = 2.73$, $SD = 1.30$), see Figure 7c.

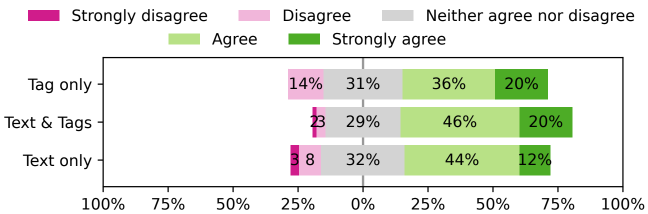
Lastly, for *Sufficient Breadth*, we see two significant differences, between *Text & Tags* and *Text-Only* ($Z = -3.607$, $p < .001$) as well as between *Tags-Only* and *Text-Only* ($Z = -3.997$, $p < .001$). In other words, **reviews that included *Text & Tags* $M = 3.53$, $SD = 1.15$) were perceived as having more breadth, i.e., spanning a larger number of different subject areas, as compared to *Text-Only* ($M = 2.71$, $SD = 1.29$). Similarly, the *Tags-Only* ($M = 3.66$, $SD = 1.15$) reviews had more breadth than the *Text-Only* reviews**. Overall, the *Tag-Only* condition received the highest average rating in this item.

5.4.3 Information Relevance. Next, we looked into participants' perceptions of the relevance of the information provided in the reviews. We break this concept down into the items *Helpful to Evaluate* (the accommodation), *Helpful to Familiarize* (oneself with the accommodation, *Easing the Decision-Making* Process, *Providing Additional Information* beyond what one would usually find in a property description, and *Reflecting Subjective Experience* of the user.

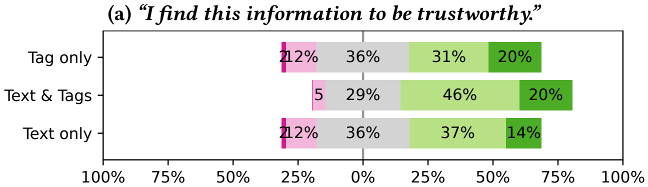
Again, none of the data in this section is normally distributed. Thus, we performed Friedman tests, which showed significant differences between the three conditions for all items except *Reflecting Subjective Experience*; see Table 1. Afterward, we performed Wilcoxon signed-rank tests with Holm–Bonferroni correction applied.

For the item *Helpful to Evaluate*, we found significant differences between *Text & Tags* and *Text-Only* ($Z = -2.917$, $p = .004$) as well as between *Tags-Only* and *Text-Only* ($Z = -2.614$, $p = .009$). In other words, **reviews that include *Text & Tags* ($M = 3.93$, $SD = .89$) were perceived as most helpful to evaluate a property**, followed by *Tags-Only* ($M = 3.90$, $SD = .904$) and lastly *Text-Only* ($M = 3.42$, $SD = 1.13$).

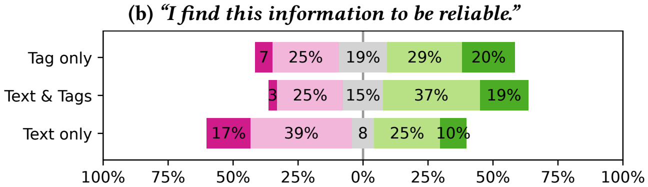
The results for *Helpful to Familiarize* (see Figure 7d) revealed that there is a statistically significant difference between *Text-Only* with both *Tags-Only* ($p < .021$) and *Text & Tags* ($p = .005$). However, not



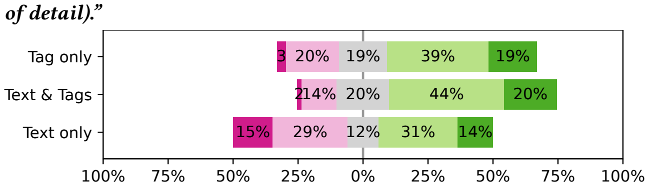
(a) *"I find this information to be trustworthy."*



(b) *"I find this information to be reliable."*



(c) *"The information from the reviews provided sufficient detail (level of detail)."*



(d) "The review provided was helpful for familiarizing me with the accommodation".

Figure 7: Comparison of the three review categories regarding users' perception of the reviews being trustworthy (a), reliable (b), providing sufficicient detail (c), and being helpful (d). The figures show the number of responses on a 5-point Likert scale from 1 (=strongly disagree) to 5 (=strongly agree).

between *Tags-Only* and *Text & Tags* (p = .275). ***Text & Tags* was perceived most helpful** (M = 3.68, SD = 1.01), outperforming the other two conditions *Tags-Only* (M = 3.49, SD = 1.12) and *Text-Only* (M = 2.98, SD = 1.33).

Moreover, we found a significant difference in *Easing Decision-Making* ($Z = -2.562$, p = .01). ***Text & Tags* reviews (**M = 3.73, SD = 1.10**) made it significantly easier to come to a decision on booking a property compared to *Text-Only* reviews** (M = 3.29, SD = 1.16).

Lastly, the item *Providing Additional Information*, the post-hoc test revealed significant differences between *Text & Tags* and *Text-Only* ($Z = -3.453$, $p <$ .001) as well as *Tags-Only* and *Text-Only* ($Z = -2.798$, p = .005). In detail, **in both conditions that include tags,** (*Text & Tags*: M = 3.59, SD = 1.19; *Tags-Only*: (M = 3.44, SD = 1.21)) **participants state that they agree that the reviews contain information beyond what is available in an accommodation description.** In the *Text-Only* conditions, participants were more neutral on this statement (M = 2.80, SD = 1.35).

5.4.4 Open Comments. In the last open feedback section, we asked participants which review format they favored and to elaborate on their decision. Overall, most participants preferred the combination of textual feedback with tags *(N=21)*. They stress, for example, that "*It was a mixture of key highlights as well as subjective feedback*" (S2P12) or that "*The person elaborated in text about the negatives of the place, which I found really helpful. They still put lots of tags in, which was all useful information*" (S2P14). The richness of information was positively emphasized; for example, S2P7 noted that "*It gave details of things I may not have thought about like limescale.*"

Other participants were more hesitant about the tags. S2P42 describes it nicely when saying that they "*[...] did like the simplicity of the tags and the rating systems, [they] do find that written reviews come across as more genuine. [...] Ratings and logos alone can sometimes feel quite 'bot like'* ". Similar opinions are voiced by other participants, saying that in the written text, they "*[...] like a personal touch, but the tags were useful in pointing out some extras*" (S2P18) or that written reviews provide more "*[...] personal experience*" (S2P33). For S2P1, it was particularly important that the review showed that the author "*[...] cared about their review [...]*" (S2P1).

Lastly, some participants critiqued the tag-based interface by saying that "*Too many icons take the focus away from the text*" (S2P14) and that "*Some of the blue tags [were] irrelevant*" (S2P54).

6 Discussion

6.1 Limitations

We referred to existing systems in designing our review prototype, which already comes with their own limitations. The choice of presenting each question on a separate page instead of a one-page layout can be debated regarding its effect on users' experience. One participant suggested the inclusion of open-ended text fields and tags for every question to enrich content generation. However, such an approach could be seen as cumbersome as it increases typing load even more, potentially lowering response rates. Furthermore, question order might influence outcomes and willingness to produce detailed reviews; starting with negatives might prime participants to view their stay more negatively.

In our prototypes we introduce different scales: a star based one and an emoji based one. As the research focus was on the effects tags have as UI elements, we did not further inspect the effects that the different scales might have caused in the review creation process. Perhaps the visual ques (star or emoji) could have had an influence on the perception and therefore the inputted rating, however further research would be needed to inspect such claims.

Furthermore, given the anonymous participant selection, we know very little about their actual stay. Depending on the time since the trip, refreshing their memories about a past stay might have been more or less challenging.

Another significant limitation was the absence of context about the specific accommodation under review in the second user study. Since we did not want to introduce potential effects caused by the quality of an accommodation, we opted not to provide any details on the actual accommodation at all. The absence of specific context for reviews might have hindered participant immersion and made it harder for our participants to actually assess the reviews, particularly their trustworthiness, as contextual factors can influence the perceived helpfulness of reviews [41]. The reviews also lacked personal touches, such as names or photos, which could affect their perceived credibility. Anonymity, while ensuring objectivity, might make reviews seem impersonal or fabricated. Still, it was necessary not to introduce another source of bias (e.g., age, gender, nationality, emotion in the picture, etc.). Ultimately, by keeping the anonymity consistent across all conditions, we could still compare users' perceptions of them.

6.2 User Experience of Tags for Review Creation

For the first user study on the creation of reviews, we followed a between-subject design. While a within-subject design would allow participants to draw direct comparisons among the interfaces, we opted for a between-subject design to limit the bias that seeing the other condition would create. In the results of our study, we see that both prototypes were rated very positively, limiting the chances of finding significant differences in the user experience ratings. Particularly, the fact that our control prototype reflects the standard review creation UI, as many bigger accommodation booking websites implement it, meant that people did not rate it poorly. However, when looking specifically at the perception of tags, users indicate that they consider them helpful and easy to use. We could not find a difference in perceived effort among the conditions. However, we did not collect data on the actual time it took users to create the reviews. Nonetheless, we can conclude that tags are perceived as overall beneficial for the review creation process.

6.3 User Experience of Tags for Review Perception

In our second user study, we evaluated users' perspectives on the three review conditions: *Tags-Only*, *Text-Only*, and *Text & Tags* combined. We see that the *Text & Tags* condition outperforms the other two conditions in almost every facet of general understanding, information quality, quantity, and relevance. However, similarly, the *Tag-Only* condition outperformed the *Text-Only* condition in almost every facet, too. Thus, we can show that Tags alone can create reviews that are equally and in some situations even more helpful, trustworthy, and of sufficient detail and information breadth than textual reviews. Especially the *Text-Only* condition was perceived as having less diversity of content (i.e., information breadth). This confirms our assumption that tags can trigger users to provide information about an accommodation that they would not have provided without tags as memory cues. Furthermore, we could not find a difference among the three conditions regarding the users' perception of reliability and trustworthiness. While we do not have specific information regarding what users think about who generated the reviews, we can speculate that the lack of freely written text does not significantly decrease the perceived humanness of the reviews. However, it is likely that these subjective perceptions are influenced by the phrasing of the actual tags.

6.4 Tag Generation

In the tag-based prototypes, we generated a list of fixed tags as potential choices from frequently mentioned phrases on booking platforms such as Airbnb and Booking.com. This list is obviously limited and will have influenced the usage of the tags in our study. A more targeted selection of tags based on property descriptions, frequent phrasings from the individual user, or generative AI could further improve the selection rate of the tags and enrich the reviews. Further investigations need to determine the preferred tag designs, such as the number of tags per rating or text field, length (i.e., individual word vs. sentence snippet), specificity, and level of detail. Striking a balance between a diverse selection of tags and a tidy and clean interface will be challenging.

6.5 Review Creation and Perception in Times of AI Tools

Over the last two decades, researchers have already explored automating review creation and perception, such as review summarization or aggregation tools or providing guiding questions. However, post-hoc experience reports can lack certain details or be skewed or biased (e.g., recency bias [22]). With the help of novel technologies and AI, we could move the review creation to the relevant moments. For example, guests could be prompted during the hotel check-in or certain moments during the day to narrate their experience. Through Natural Language Processing (NLP), we could further extract sentiments or provide personalized meta-reviews to fit individual tastes and preferences. Yet, we expect that peer-generated reviews will remain essential. While AI can provide efficiency, achieving similar levels of personal and emotional connection as well as trustworthiness will be difficult. Furthermore, aggregating technologies risk losing extreme reviews and lowering the diversity of opinions.

7 Conclusion

This work examined the influence of tags in review creation and perception for an accommodation booking scenario. We first assessed the UI of platforms like Airbnb and Booking.com. Noting that existing systems often require extensive text input or lack details on scale input, we prototyped a combined review system introducing "tags" with traditional elements. This concept was tested in two user studies. In the first study, we asked participants to create a review using our interface. Participants found tags clear and effective for expressing opinions. Our participants found them easily comprehensible and useful in expressing opinions without increasing the perceived effort. The combination of text fields and tags allowed for well-balanced reviews, while pairing rating scales with tags provided clearer explanations for scores, e.g., for attributes like "Cleanliness."

The second study highlighted tags' value from the review perception perspective, where we introduced tags as an additional source of information in the presented user reviews. Reviews featuring tags showed better detail and quality. Tags revealed overlooked accommodation specifics, such as cleanliness or amenities, bolstering consumer decisions. Among the three review types studied (*Tags-Only*, *Text & Tags*, and *Text-Only*), we suggest integrating text and tags. Tags ease information input and minimize typing for users. However, Tags-Only was perceived very well overall, so it can serve as a valuable option for gathering quick reviews in mobile settings. In conclusion, our research positions tags as essential UI elements, refining the review process and elevating review quality. They guide users in producing in-depth, factual reviews, facilitating more informed consumer decisions.

Open Science

We encourage readers to reproduce and extend our results. Therefore, we made the data collected in our study and our analysis scripts available on the Open Science Framework https://osf.io/34cjf/.

References

[1] Eytan Adar, Jaime Teevan, and Susan T. Dumais. 2008. Large scale analysis of web revisitation patterns *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1197–1206. https://doi.org/10.1145/1357054.1357241

[2] Sarah Alismail and Hengwei Zhang. 2018. The use of emoji in electronic user experience questionnaire: An exploratory case study. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. Hawaii, Honolulu, HI, USA, 3366–3375. http://hdl.handle.net/10125/50315

[3] Latha Banda and K.K. Bharadwaj. 2014. An Approach to Enhance the Quality of Recommendation Using Collaborative Tagging. *International Journal of Computational Intelligence Systems* 7, 4 (2014), 650–659. https://doi.org/10.1080/18756891.2014.960225

[4] Inès Blal and Michael C Sturman. 2014. The differential effects of the quality and quantity of online reviews on hotel room sales. *Cornell Hospitality Quarterly* 55, 4 (2014), 365–375. https://doi.org/10.1177/1938965514533419

[5] Judith Bridges and Camilla Vásquez. 2018. If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism* 21, 18 (2018), 2057–2075. https://doi.org/10.1080/13683500.2016.1267113

[6] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).

[7] Frederik Brudy, Suppachai Suwanwatcharachat, Wenyu Zhang, Steven Houben, and Nicolai Marquardt. 2018. EagleView: A Video Analysis Tool for Visualising and Querying Spatial Interactions of People and Devices. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (Tokyo, Japan) *(ISS '18)*. Association for Computing Machinery, New York, NY, USA, 61–72. https://doi.org/10.1145/3279778.3279795

[8] Dimitrios Buhalis and Rob Law. 2008. Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism management* 29, 4 (2008), 609–623. https://doi.org/10.1016/j.tourman.2008.01.005

[9] Man Yee Cheung, Chuan Luo, Choon Ling Sia, and Huaping Chen. 2009. Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International journal of electronic commerce* 13, 4 (2009), 9–38. https://doi.org/10.2753/JEC1086-4415130402

[10] Gilbert A Churchill, Jr. 1979. A paradigm for developing better measures of marketing constructs. *Journal of marketing research* 16, 1 (1979), 64–73. https://doi.org/10.1177/002224377901600110

[11] Dimitar Dimitrov, Denis Helic, and Markus Strohmaier. 2018. *Tag-Based Navigation and Visualization.* Springer International Publishing, Cham, 181–212. https://doi.org/10.1007/978-3-319-90092-6_6

[12] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191. https://doi.org/10.3758/BF03193146

[13] Raffaele Filieri. 2015. What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of business research* 68, 6 (2015), 1261–1270. https://doi.org/10.1016/j.jbusres.2014.11.006

[14] Raffaele Filieri, Salma Alguezaui, and Fraser McLeay. 2015. Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism management* 51 (2015), 174–185. https://doi.org/10.1016/j.tourman.2015.05.007

[15] Diana Gavilan, Maria Avello, and Gema Martinez-Navarro. 2018. The influence of online ratings and reviews on hotel booking consideration. *Tourism Management* 66 (2018), 53–61. https://doi.org/10.1016/j.tourman.2017.10.018

[16] Duarte Gonçalves, Rui Jesus, and Nuno Correia. 2008. A Gesture Based Game for Image Tagging. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2685–2690. https://doi.org/10.1145/1358628.1358745

[17] Daniel Kailer, Peter Mandl, and Alexander Schill. 2013. Rated Tags: Adding rating capability to collaborative tagging. In *2013 International Conference on Cloud and Green Computing* (New York, New York, USA). IEEE, 249–255. https://doi.org/10.1109/CGC.2013.46

[18] Daniel Kailer, Peter Mandl, and Alexander Schill. 2014. Supporting Customers' Decision Making with Rated Tags. In *Proceedings of the Sixteenth International Conference on Electronic Commerce* (Philadelphia, PA, USA) *(ICEC '14)*. Association for Computing Machinery, New York, NY, USA, 33–40. https://doi.org/10.1145/2617848.2617849

[19] Stephen W Litvin, Ronald E Goldsmith, and Bing Pan. 2008. Electronic word-of-mouth in hospitality and tourism management. *Tourism management* 29, 3 (2008), 458–468. https://doi.org/10.1016/j.tourman.2007.05.011

[20] Maliha Mahbub, Najia Manjur, and Julita Vassileva. 2021. Towards better rating scale design: An experimental analysis of the influence of user preference and visual cues on user response. In *International Conference on Persuasive Technology*. Springer, Cham, 150–163. https://doi.org/10.1007/978-3-030-79460-6_12

[21] G A Miller. 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63, 2 (March 1956), 81–97.

[22] Bennet B Murdock Jr. 1962. The serial position effect of free recall. *Journal of experimental psychology* 64, 5 (1962), 482. https://doi.org/10.1037/h0045106

[23] Beng Soo Ong. 2012. The perceived influence of user reviews in the hospitality industry. *Journal of Hospitality Marketing & Management* 21, 5 (2012), 463–485. https://doi.org/10.1080/19368623.2012.626743

[24] Paul Phillips, Stuart Barnes, Krystin Zigan, and Roland Schegg. 2017. Understanding the impact of online reviews on hotel performance: an empirical analysis. *Journal of Travel Research* 56, 2 (2017), 235–249. https://doi.org/10.1177/0047287516636481

[25] Rajinesh Ravendran, Ian MacColl, and Michael Docherty. 2012. A Comparative Usability Study of a Tag-Based Interface in the Mobile Banking Context. In *Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction* (Dunedin, New Zealand) *(CHINZ '12)*. Association for Computing Machinery, New York, NY, USA, 80–83. https://doi.org/10.1145/2379256.2379269

[26] Rajinesh Ravendran, Ian MacColl, and Michael Docherty. 2012. Tag-based interaction in online and mobile banking: a preliminary study of the effect on usability. In *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction* (Matsue-city, Shimane, Japan) *(APCHI '12)*. Association for Computing Machinery, New York, NY, USA, 35–40. https://doi.org/10.1145/2350046.2350056

[27] Rajinesh Ravendran, Ian MacColl, and Michael Docherty. 2012. Usability Evaluation of a Tag-Based Interface. *J. Usability Studies* 7, 4 (aug 2012), 143–160.

[28] Simon André Scherr, Svenja Polst, Lisa Müller, Konstantin Holl, and Frank Elberzhager. 2019. The Perception of Emojis for Analyzing App Feedback. *International Journal of Interactive Mobile Technologies* 13, 2 (2019), 18. https://www.learntechlib.org/p/207545/

[29] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2014. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, Aaron Marcus (Ed.). Springer International Publishing, Cham, 383–392.

[30] Beverley A Sparks and Victoria Browning. 2011. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management* 32, 6 (2011), 1310–1323. https://doi.org/10.1016/j.tourman.2010.12.011

[31] Florian Spiess, Ralph Gasser, Heiko Schuldt, and Luca Rossetto. 2023. The Best of Both Worlds: Lifelog Retrieval with a Desktop-Virtual Reality Hybrid System. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge*. Association for Computing Machinery, New York, NY, USA, 65–68. https://doi.org/10.1145/3592573.3593107

[32] Vera Toepoel, Brenda Vermeeren, and Baran Metin. 2019. Smileys, Stars, Hearts, Buttons, Tiles or Grids: Influence of Response Format on Substantive Response, Questionnaire Experience and Response Time. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 142, 1 (2019), 57–74. https://doi.org/10.1177/0759106319834665

[33] Alexander Toet and Jan BF van Erp. 2019. The EmojiGrid as a tool to assess experienced and perceived emotions. *Psych* 1, 1 (2019), 469–481. https://doi.org/10.3390/psych1010036

[34] Christoph Trattner, Yi-ling Lin, Denis Parra, Zhen Yue, William Real, and Peter Brusilovsky. 2012. Evaluating Tag-Based Information Access in Image Collections. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media* (Milwaukee, Wisconsin, USA) *(HT '12)*. Association for Computing Machinery, New York, NY, USA, 113–122. https://doi.org/10.1145/2309996.2310016

[35] Shawn Tseng and B. J. Fogg. 1999. Credibility and Computing Technology. *Commun. ACM* 42, 5 (may 1999), 39–44. https://doi.org/10.1145/301353.301402

[36] Ivar E Vermeulen and Daphne Seegers. 2009. Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management* 30, 1 (2009), 123–127. https://doi.org/10.1016/j.tourman.2008.04.008

[37] Jesse Vig, Matthew Soukup, Shilad Sen, and John Riedl. 2010. Tag Expression: Tagging with Feeling. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 323–332. https://doi.org/10.1145/1866029.1866079

[38] Richard Y Wang and Diane M Strong. 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems* 12, 4 (1996), 5–33. https://doi.org/10.1080/07421222.1996.11518099

[39] Sung-Byung Yang, Seung-Hun Shin, Youhee Joun, and Chulmo Koo. 2017. Exploring the comparative importance of online hotel reviews' heuristic attributes in review helpfulness: a conjoint analysis approach. *Journal of Travel & Tourism Marketing* 34, 7 (2017), 963–985. https://doi.org/10.1080/10548408.2016.1251872

[40] Qiang Ye, Rob Law, and Bin Gu. 2009. The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management* 28, 1 (2009), 180–182. https://doi.org/10.1016/j.ijhm.2008.06.011

[41] Xiaobin Zhang and Hak-Seon Kim. 2021. Customer experience and satisfaction of Disneyland hotel through big data analysis of online customer reviews. *Sustainability* 13, 22 (2021), 12699. https://doi.org/10.3390/su132212699