CrowdSurfer: Seamlessly Integrating Crowd-Feedback Tasks into Everyday Internet Surfing

Saskia Haug
saskia.haug@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Ivo Benke
ivo.benke@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Daniel Fischer
uowcd@student.kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Alexander Maedche
alexander.maedche@kit.edu
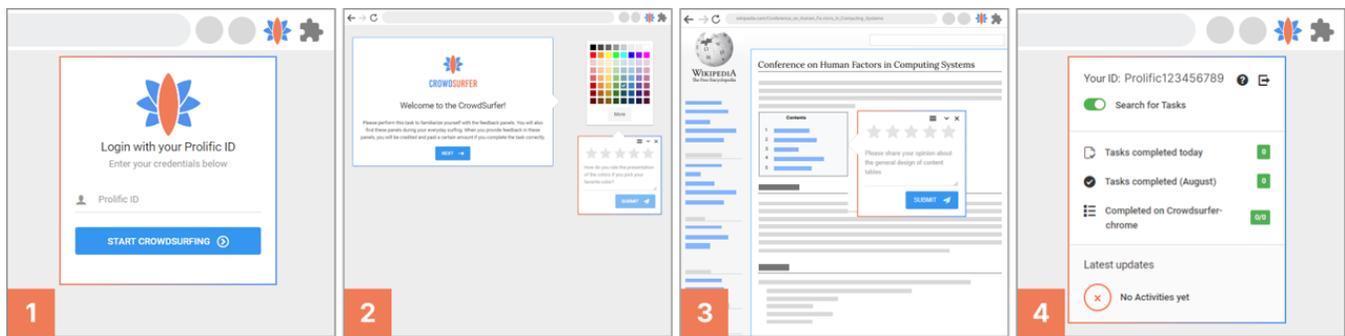Karlsruhe Institute of Technology
Karlsruhe, Germany

Figure 1: The CrowdSurfer extension: 1) Crowdworkers install the CrowdSurfer and register with their ProlificID. 2) The CrowdSurfer is explained in a demo task. 3) Crowdworkers can solve feedback tasks during everyday internet surfing. 4) Crowdworkers can manage tasks and payments via the CrowdSurfer extension.

ABSTRACT

Crowd feedback overcomes scalability issues of feedback collection on interactive website designs. However, collecting feedback on crowdsourcing platforms decouples the feedback provider from the context of use. This creates more effort for crowdworkers to immerse into such context in crowdsourcing tasks. In this paper, we present CrowdSurfer, a browser extension that seamlessly integrates design feedback collection in crowdworkers' everyday internet surfing. This enables the scalable collection of in situ feedback and, in parallel, allows crowdworkers to flexibly integrate their work into their daily activities. In a field study, we compare the CrowdSurfer against traditional feedback collection. Our qualitative and quantitative results reveal that, while in situ feedback with the CrowdSurfer is not necessarily better, crowdworkers appreciate the effortless, enjoyable, and innovative method to conduct feedback tasks. We contribute with our findings on in situ feedback collection

and provide recommendations for the integration of crowdworking tasks in everyday internet surfing.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; • **Software and its engineering** → **Designing software**.

KEYWORDS

Crowd-feedback system, crowdsourcing, design feedback, browser extension

1 INTRODUCTION

The continuous evaluation of interactive designs with users is crucial for the acceptance of and user satisfaction with interactive systems [25, 33]. For example, effective evaluation techniques have been recognized as essential for websites in order to successfully attract customers [8]. A typical means for effective evaluation is the collection of user feedback in situ, during the usage of a website. In

situ feedback collection using pop-ups or feedback buttons is a powerful way to identify problems, critically reflect on existing features, or collect new additional requirements to increase users' acceptance [43]. However, users often perceive such feedback requests as hindering and annoying. In general, the willingness to engage with the feedback request and share meaningful feedback is low [2]. A potential solution to counteract this engagement challenge is the inclusion of paid crowdworkers to gather design feedback, also called crowd feedback. Crowd-feedback systems allow the large-scale collection of feedback via crowdsourcing tasks on platforms like Amazon Mechanical Turk (MTurk) and Prolific [21]. Crowd feedback has shown to be a scalable approach for successfully collecting diverse opinions and improving interactive designs [31, 35, 47].

Against these benefits, crowd feedback has major drawbacks: First, crowdworkers lack the actual context of use when providing feedback on interactive systems like websites [43]. They are most likely not real users of the respective website and do not really experience it. Potentially, they are not even familiar with the specific context of the website. This might distort their feedback. Offering crowdworkers context has shown to increase their empathy [4] and, in turn, to improve the feedback quality and quantity [47]. Context, e.g., in the form of personas has a positive impact on empathy because this helps to recognize and understand the real users' thoughts and feelings [10]. Second, if crowd-feedback systems offer an artificial usage context like a scenario or a persona (e.g., [22]) time and effort for workers to immerse in the artificial usage scenario increase. This additional effort must, of course, also be compensated. A mismatch between the required time and effort for a task and the monetary reward is the main reason for crowdworkers to return, abandon, or reject tasks and is also one of the two causes for crowdworkers' poor hourly wages of around $2 to $5 [19, 26]. Third, crowd-feedback studies on crowdsourcing platforms usually run at a specific point in time. They collect feedback only on a snapshot of the system and do not allow to collect feedback continuously. Due to these drawbacks, in situ feedback is preferable since the feedback providers experience the system and its functionalities in context and real-time [32, 36, 42]. However, there is still a lack of knowledge on the differences between in situ feedback and feedback that is collected separately from the actual usage (e.g., in a survey-based crowdworking task), especially regarding feedback quality and quantity. Current research lacks an approach that tackles these three drawbacks of crowd feedback simultaneously. The main focus of research on crowd feedback is still to maximize the feedback quality and quantity. The crowdworker perspective is often neglected [20]. We assume it may be promising to ask crowdworkers for feedback in situ when they are actual users of the system. Simultaneously, this decreases the additional effort of crowdworkers to immerse into feedback tasks, thereby reduces the additional hidden or invisible work for searching, selecting, and accepting the task, and makes the payment fairer. Allowing crowdworkers to solve tasks during their everyday internet surfing might also increase the flexibility of their working conditions. Ergo, our goal is to integrate feedback tasks into their everyday internet surfing. Thus, we leverage crowdworkers as real users and empower them to work during internet usage. With our field study, we aim to understand how crowdsourced real and in situ user feedback differs from traditional crowd feedback and how crowdworkers perceive the integration of crowdworking tasks into their everyday internet surfing.

Following this objective, in this paper, we present *CrowdSurfer*, an innovative crowd-feedback system in the form of a browser extension that allows crowdworkers to provide website design feedback during their everyday internet surfing in return for a monetary reward. Thereby, we combine the benefits of crowdwork and traditional user feedback. Functionally, the CrowdSurfer is connected to a crowdsourcing platform. After installation, crowdworkers can work on existing tasks when visiting respective websites during their everyday internet surfing. Figure 1 shows the process of using the CrowdSurfer extension as a crowdworker.

We evaluated the CrowdSurfer in an experimental field study with 63 crowdworkers following a quantitative and qualitative approach. We assessed the feedback quality and quantity of the design feedback provided as well as the perceptions of the crowdworkers regarding the usability of the CrowdSurfer, the feedback process, and its effect on the working conditions of crowdworkers. Our results show that the CrowdSurfer was comfortable, simple, and enjoyable for the crowdworkers to use and that they perceived conducting feedback tasks with the CrowdSurfer as fairer regarding payment and effort. Although participants stated that they believe the feedback they provided with the CrowdSurfer is more real and therefore more relevant, quantitative results showed that the CrowdSurfer feedback is less specific, actionable, relevant, and shorter. This discrepancy enlightens an interesting differential between positive effects on the working conditions of crowdworkers and a lower quality of design feedback comments in comparison with traditional feedback tasks. We identified important aspects that demonstrate the utility of the CrowdSurfer for requesters despite the reduced feedback quality. Based on our findings we derived recommendations for the future design of crowdsourcing systems integrated into crowdworkers' everyday internet surfing. With our work, we contribute:

- The *CrowdSurfer*, a crowd-feedback system in form of a Chrome extension for crowdsourcing feedback on websites in return for monetary rewards.
- Findings demonstrating the utility of a browser extension to include feedback tasks in crowdworkers' everyday internet surfing (e.g., with regards to the effort of work, fairness of payment, and flexibility).
- Design recommendations for developing future crowdsourcing systems that integrate tasks into crowdworkers' everyday internet surfing.

2 RELATED WORK

In the following, we present related work on crowdsourcing design feedback, casual microtasking and in situ feedback, and crowdworker working types, behaviors, and conditions.

2.1 Crowdsourcing Software Design Feedback

User involvement in the continuous evaluation of website designs is crucial. Traditionally, websites are evaluated using methods like usability tests, interviews, or focus groups [46]. As these methods lack scalability, are costly, and require access to users, crowd feedback has evolved as a complementary approach for collecting

large amounts of design feedback. Crowd feedback comes from the visual design domain where feedback is usually provided by peers [51]. Various crowd-feedback systems have been suggested to collect quantitative and qualitative design feedback for websites [22, 35] or mobile apps [42]. Further, such crowd-feedback systems often include numerous design features to enrich the feedback [21]. Research has shown that crowd-feedback systems are capable of achieving a feedback quality similar to expert feedback [31]. One of the first crowd-feedback systems is *Voyant* [50]. Voyant was designed to collect feedback on poster designs by collecting impressions of the crowd and analyzing the adherence to design guidelines. Voyant combined the collection of qualitative and quantitative feedback with a marker feature so that feedback providers could draw boxes to highlight a designated area and support their textual feedback [50].

While some systems similar to Voyant are focused on feedback collection during the development process (e.g., [34, 39, 47]), others collect feedback after go-live for continuous improvement (e.g., [35, 42, 44]). While crowd-feedback systems provide multiple benefits to continuously evaluate the software designs during the development process, they have downsides as well. Often, they only allow feedback collection in dedicated studies and continuous in situ feedback collection is not possible. Furthermore, crowdworkers do not actually use the software and the feedback is provided in an artificial usage context, for example, with a persona as context [5], or a usage scenario to consider when providing feedback [22]. Our study extends prior work by offering an innovative way to crowdsource website design feedback. With the CrowdSurfer we tackle the mentioned problems by enabling crowdworkers to conduct design feedback tasks during their everyday internet surfing. Additionally, we want to contribute with a better understanding of the impact of the context on feedback quality.

2.2 Integration of Feedback Tasks in Internet Surfing

Hahn et al. [16] invented the term *casual microtasking* to describe the integration of microtasks into other primary activities of workers. In their study, they inserted writing microtasks into the Facebook feed to allow workers to solve microtasks during short breaks. Their results indicate that *casual microtasking* is a promising approach to leveraging spare micromoments [16]. Further studies investigated the role of the context of crowdworkers when accepting tasks. The results of Goncalves et al. [14] highlight the potential of context to motivate participation in ubiquitous crowdsourcing tasks. They showed that if the crowdsourcing task is located directly next to the physical element on which feedback is collected the participation rate increases. Therefore, situatedness in feedback tasks seems to increase participation rates and engagement. Also, the crowdworker context influences task acceptance and crowdworker preferences [23].

The integration of feedback tasks into everyday internet surfing leads to the collection of so-called in situ feedback. In situ feedback is user feedback that is collected while the user is actually using and experiencing the system. A key advantage of in situ feedback is that users do not have to leave the experience to provide feedback which means less interruption to them [36]. There exist dedicated

systems to collect in situ feedback. *AppEcho* [42] is a mobile feedback approach that allows users to provide feedback about their smartphone applications. *iRequire* [41] is a similar system that allows users to provide feedback on their environment, such as a timetable at a bus stop. In the application, they can take a picture and add a textual description of their related requirements. In situ feedback may also be combined with passive logging data as applied by *MyExperience* [12], a system that captures device usage, user context, and environmental sensing in the background. Additionally, *MyExperience* conducts user experience sampling to collect in situ user feedback.

These studies have demonstrated the feasibility and advantages of capturing in situ feedback. However, existing research has mainly focused on developing mobile applications to capture in situ feedback. In our study, we want to provide a crowd-feedback system for collecting in situ feedback from crowdworkers. To do this, we want to leverage the approach of casual microtasking based on the results of Hahn et al. [16], Goncalves et al. [14], and Hettiachchi et al. [23] by providing further insights on how to integrate tasks into crowdworkers' daily life. Thereby, we aim to understand how in situ feedback differs from traditional survey-based feedback.

2.3 Working Conditions of Crowdworkers

Crowd work is a well-researched topic in the field of human-computer interaction. When designing an innovative approach for crowdsourcing tasks, we need to understand the crowdworkers' characteristics, working behavior, and preferences as well as their problems, requirements, and restrictions. This allows for informing the design rationales for our CrowdSurfer extension.

Research on crowdworkers' characteristics showed that many of them are multitaskers and mix work and non-work activities [30]. This finding is also supported by Williams et al. [49] who found out that crowdworkers tend to divide their attention between work and non-work related activities (e.g., watching TV). This may be partly caused by the support tools frequently used by crowdworkers. These tools (e.g., MTurk Suite [1], TurkerView [2]) enable and reinforce task-switching and multitasking behavior. They also promote the fragmentation of crowdworkers' work-life boundaries as they enable a 'work-anywhere' attitude [49]. To better understand the work practice of crowdworkers, Williams et al. [49] also investigated the work-life boundaries of crowdworkers. In their study, the majority of participants had a low boundary control, meaning they felt they cannot control the timing, frequency, and direction of boundary crossings regarding interruptions to fit their identities [29].

Crowdworking platforms (e.g., MTurk, Prolific, CrowdFlower) differ in the types of users they attract. For example, in the study of Abbas and Gadiraju 41% of participants on MTurk reported using MTurk as their main source of income, while only 8% of Prolific users reported the same for Prolific [1]. This is consistent with the results of earlier studies [6] and shows that Prolific workers are potentially more open to casual microtasking as they are not purely focusing on maximizing their financial rewards. Also, the social

---

protection and working conditions of crowdworkers have already been investigated in multiple studies (e.g., [9, 11, 30]). Frequently mentioned problems regarding the working conditions are the limited flexibility of crowdworkers [30] and the low payment which is partly caused by invisible work [18]. The term invisible work summarizes unpaid but necessary duties of crowdworkers such as job search, task rejection, task submission, and task information gathering [18, 38]. To address the mentioned challenges, recent research already proposed extensions for crowdworkers to better manage their tasks, increase transparency, and give crowdworkers a voice (e.g., *TurkScanner* [38], *Turkopticon* [24], and *Turker Tales* [27]).

We believe that it is important for our crowd-feedback system to consider crowdworkers' characteristics and enable the conduction of fair crowdworking tasks. This includes supporting crowdworkers in setting boundaries between work and non-work related activities, counteracting invisible work, and increasing flexibility.

3 THE CROWD-FEEDBACK SYSTEM CROWDSURFER

The goal of our design solution is twofold. First, we want to combine the benefits of user feedback, especially the real context, with the scalability of crowdsourcing for design feedback collection. Second, we want to improve the working conditions for crowdworkers and provide them with a fair and flexible way of working. Therefore, we decided to design a crowd-feedback system in form of a browser extension that enables crowdworkers to gain monetary rewards during their everyday internet surfing.

Before developing a full-functioning browser extension, we first developed an initial prototype, which we then discussed in exploratory interviews with five crowdworkers (four female, one male). The participants were recruited on Prolific and were on average 34.80 years old (SD = 10.76) with one to seven years of crowdworking experience. These interviews helped us to elaborate our design rationale and understand how distinctive features need to be implemented in the final CrowdSurfer system.

3.1 Design Rationale

With respect to the goals of our study, the design of our crowd-feedback system follows three fundamental design rationales:

(1) *Seamless integration in everyday internet surfing.* The main goal of our crowd-feedback system is to allow seamless integration of design feedback tasks in crowdworkers' everyday internet surfing. Crowdworkers shall not be distracted from their primary tasks but still notice the availability of feedback tasks. To achieve high adoption, it is crucial that users do not get annoyed by feedback requests.

(2) *Control for crowdworkers.* As our feedback extension impacts crowdworkers during their everyday internet surfing, they need to be in control over the system in general and the tasks in particular. They also need to be able to control their boundaries between non-work and work activities.

(3) *Feedback value.* The system needs to generate high-quality feedback to present value to feedback requesters. The system shall be able to collect different types of feedback to address requesters' needs.

3.2 System Design

In this chapter, we will present the final design of the CrowdSurfer and the implemented features. We explain our design decisions by referring to related work or to our exploratory interviews. To allow for seamless integration into crowdworkers everyday internet surfing, we decided to implement the crowd-feedback system as a Chrome extension. This Chrome extension displays feedback tasks as pop-ups on the respective websites. In the following, we describe the design and features of the CrowdSurfer according to three steps: (1) Download and setup, (2) providing feedback, and (3) managing tasks.

3.2.1 Download and Setup. The CrowdSurfer can be installed via the Chrome web store. Crowdworkers can log in by entering their crowdsourcing platform ID (here Prolific) (see Figure 2 top). This is required so that their task submissions can be matched to and paid via their crowdwork account. Then, crowdworkers need to conduct a demo task to learn about the features of the CrowdSurfer (see Figure 2 bottom).
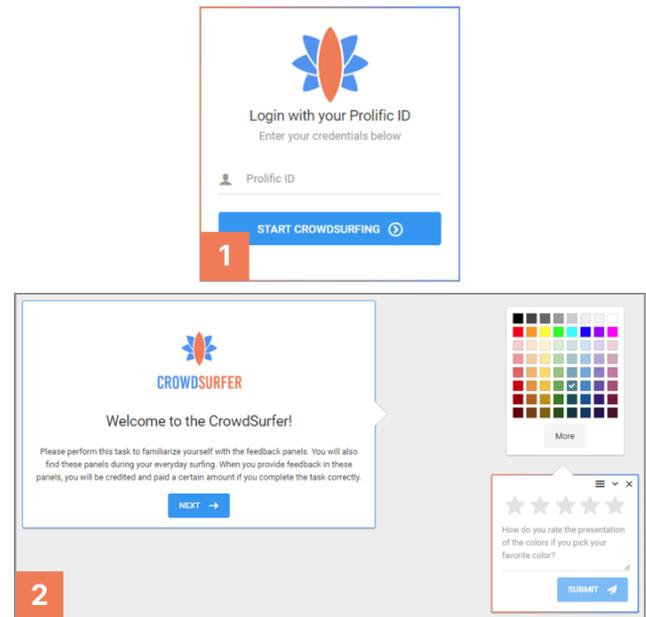
Figure 2: Setup of the CrowdSurfer: 1) Login screen, 2) demo task to explain features

3.2.2 Providing Feedback. After the setup and the demo task are completed, the CrowdSurfer will display feedback tasks as pop-ups on selected websites. A screenshot of such a feedback pop-up is displayed in Figure 3 (left). Each feedback pop-up is attached to a website element (1). The evaluation of website elements instead of a whole website, in general, allows feedback requesters to get more specific and structured feedback from users. Additionally, crowdworkers feel that providing feedback is easier when they can focus it on a specific element [20]. In our interviews, participants initially complained that "*it's not clear if I have to rate the whole area or just a part*" (I3). Consequently, the design of the feedback

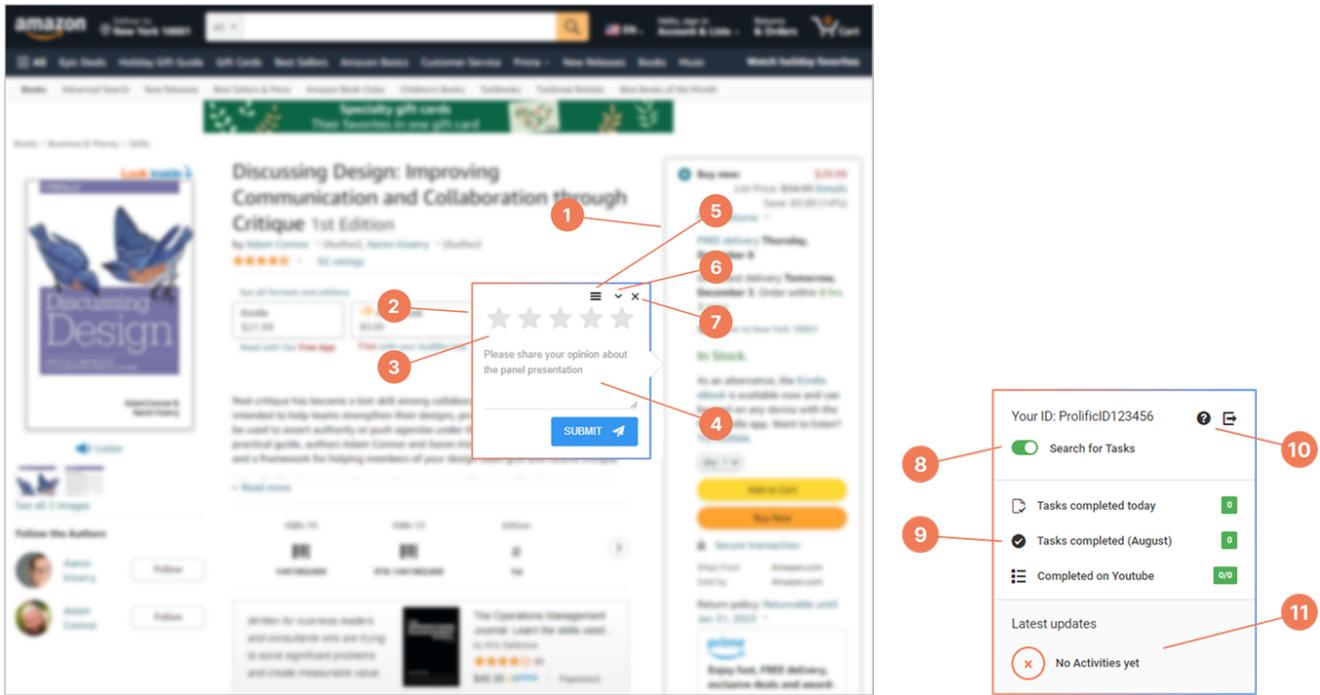Figure 3: Feedback pop-up on blurred Amazon website (left) and CrowdSurfer panel (right): 1) Element on which the feedback is collected, 2) feedback request pop-up, 3) star rating, 4) feedback text field with a question, 5) menu icon to see background information and set a reminder, 6) minimize icon, 7) reject icon, 8) toggle button to turn the CrowdSurfer on and off, 9) information on task rewards, 10) support icon to redo the demo task, and 11) overview of recently submitted tasks.

pop-ups was refined so that they clearly highlight a website element. When collecting crowd feedback from real users, usually both qualitative and quantitative feedback is collected [21]. Both feedback types have their advantages and disadvantages. Feedback providers often prefer multiple choice or ratings as it is simpler and faster [2], but qualitative text feedback of course contains more information. Therefore, our feedback extension is able to collect both types of feedback. The quantitative feedback is collected in form of a star rating (3) and the qualitative feedback as an answer to a question about the respective element (4). Each feedback pop-up has a menu (5) that allows crowdworkers to access task information of a task, e.g., the payment, the requester name, or contact information. Thereby, we want to counteract the information imbalance between feedback requesters and crowdworkers. While feedback requesters can access a lot of information about the crowdworkers, such as qualification, location, or experience, crowdworkers usually can only access limited information such as creation date and reward amount [26]. The menu allows workers to set a reminder for the task, in case they want to postpone it. Postponing of tasks would for example be helpful if *"you find a task and you realize that it's going to take longer than what you thought it would and you'd like to go back to it and finish it later"* (I1). This feature also provides workers more control over how they want to do their work. Next to the menu is a button to minimize the task (6) to allow a seamless integration into the website. The pop-up could hide important elements of the website and crowdworkers shall not be forced to

complete or reject the task just to be able to see the whole website. Finally, the cross icon (7) allows crowdworkers to reject tasks if they are not interested in solving them. After submitting a task, crowdworkers can see how many tasks on the website they solved and how many are remaining (e.g., "1/2 tasks completed").

3.2.3 Managing Tasks. When clicking on the icon of the CrowdSurfer in the list of extensions, crowdworkers can access a pop-up to manage their tasks and the CrowdSurfer extension. This pop-up is displayed in Figure 3 (right). To give crowdworkers control over the extension and their work-life boundaries, they can turn it off when they do not want to see any feedback tasks (8). In our interviews workers stated that there are situations in which they do not want to be interrupted by such an extension (*"If I'm in an interview and then I keep on being distracted by this thing that keeps popping up, then it won't work. So I would like to be able to turn it off"* (I1)). This feature also addresses crowdworkers concerns regarding their data privacy as they were worried about the extension always tracking their online behavior and data. Next, the pop-up should contain an overview of the number of completed tasks so that crowdworkers *"see whether or not it's worth your time"* (I1). As P4 stated that they *"prefer it showing more of their earnings as opposed to how many tasks you sold because the earnings can help you dictate how much you're going to earn in total, sort of a certain target"*, we do not show the number of tasks but the total reward for the day and the month (9). Additionally, the panel shows how many tasks are open on the current website, so that crowdworkers do not start searching for

tasks when there are none. In case crowdworkers want to redo the demo task to learn about the CrowdSurfer's functionalities again, they can do this by clicking on the question mark icon (10). Finally, crowdworkers requested in our interviews to somehow be able to see what the last task was that they submitted because "*in an instance where you had a network issue [...] that latest update is going to be helpful for you to go in and see if you really have completed the task*" (I5). We show this information in the latest updates (11).

4 EVALUATION STUDY

To analyze the effects of the CrowdSurfer we conducted a quantitative and qualitative field study with crowdworkers on Prolific. The goal of this field study was to understand the benefits of Crowd-Surfer feedback compared to traditional survey-based design feedback in terms of feedback quality and quantity. Further, we aimed to understand the impact of the CrowdSurfer on the working conditions of crowdworkers such as effort of tasks, fairness of payment, and flexibility.

4.1 Procedure

For the evaluation, we decided to implement in total 13 Crowd-Surfer tasks on eight of the most frequently used websites (YouTube, Amazon, Twitter, Wikipedia, eBay, CNN, Weather.com, and Reddit) to ensure that participants will visit the websites coincidentally. Participants had seven days to use the CrowdSurfer and provide feedback on the respective websites. For each star rating, they received £0.03, and for each text feedback £0.12. After seven days, participants were notified that they can now participate in a post-task questionnaire on Prolific. This questionnaire also offered them the option to schedule a 20-minutes interview with us in return for a £4 bonus payment. For the baseline treatment, we developed a simple feedback task with a questionnaire that showed links and screenshots of websites and asked for feedback on specific elements (see Figure 4). The payment per feedback was the same as in the CrowdSurfer treatment. Afterward, they also received the same post-task questionnaire. In the post-task questionnaires, we included several attention checks. The websites and tasks were in both treatments the same and participants could in both treatments freely choose if they want to provide feedback or not. We performed twelve semi-structured qualitative interviews with participants of the CrowdSurfer treatment to understand how crowdworkers perceived the extension and to interpret the quantitative results. We focused in our interviews on three aspects: the usability of the CrowdSurfer, the crowdworkers' feedback process, and the impact of the CrowdSurfer on the working conditions of crowdworkers including their motivation. The study was approved by the German Association for Experimental Economic Research (GfeW).

4.2 Participants

We recruited 80 participants (30 baseline, 50 CrowdSurfer). We filtered for English native speakers to achieve a comparable feedback quality between the participants. In the CrowdSurfer condition, nine participants downloaded the CrowdSurfer but never submitted a task which resulted in 71 participants. Of these 71 participants, eight participants failed the attention checks in the questionnaire.
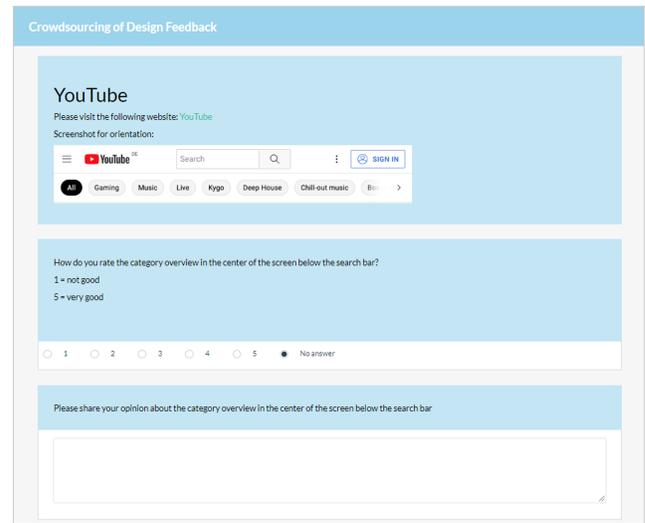

Figure 4: The design feedback survey that we used in our baseline condition.

So, we ended up with 63 final participants in our sample (29 baseline, 34 CrowdSurfer). 36.5% of the participants were female (63.5% male) and their average age was 35.8 years (SD = 12.92). Overall, the participants reported a medium experience in providing design feedback on a seven-point Likert scale (M = 3.09; SD = 1.21). They were randomly assigned to the experimental conditions.

4.3 Data Collection & Analysis

Quantitative data analysis. We gathered data in two ways. First, via the post-task questionnaire, we collected data on the perceived task duration, the perceived fairness of payment, and the perceived flexibility. For the perceived fairness of payment, we reused the items of Schulze et al. [40] that were previously used to measure the fairness in pay [3]. For the perceived flexibility we adapted the items of Kokoç [28] and Richman et al. [37].

Second, we analyzed crowdworkers' interaction with the Crowd-Surfer. We tracked how many tasks were shown and how they interacted with them. We logged when they interacted with one of the features of the CrowdSurfer (reminder, minimize task, reject task, show task information, turn the search for tasks on/off). Additionally, we collected and analyzed the feedback that crowdworkers gave in both treatments.

Feedback quality evaluation. To analyze the quality of the collected feedback comments, we conducted a separate crowdsourcing task in which UI design experts assessed the quality of the design feedback comments in six dimensions. For this task, we again used Prolific where we were able to filter for crowdworkers with UI design skills by using the respective filter. Thereby, we recruited 103 crowdworkers with experience in UI design (M = 4.51, SD = 1.53, self-assessed on a seven-point Likert scale) for the quality assessment.

To provide feedback evaluators context, we sorted the feedback comments according to the website and element they belong to. For

each website, we created separate tasks that showed each crowdworker on which element the comment was provided and presented them with up to 20 feedback comments to assess. Each feedback comment of both treatments was analyzed by three participants on the following dimensions: specificity, explanatory, actionable, positivity, relevance, and overall feedback quality. The dimensions were adopted from the study of Oppenlaender et al. [34]. The feedback quality value for each construct was assessed by taking the average from the distinct ratings of the three individual crowdworkers.

Qualitative data analysis. The interviews were transcribed and analyzed by two of the authors. We analyzed the feedback through a deductive thematic analysis following Braun and Clarke [7] based on the main topics of our interviews: CrowdSurfer usability, the feedback process, and the working conditions of crowdworkers. After the deductive analysis, we inductively refined the coding scheme. Finally, all interviews were coded by two authors. Disagreements were discussed until a consensus was found.

5 RESULTS

To investigate the effect of the CrowdSurfer on the design feedback and the crowdworker experience we conducted a three-folded analysis. First, we present the usage behavior of crowdworkers with the CrowdSurfer based on the log data. Second, we present results on crowdworkers' perceptions regarding the perceived time they spent working on tasks, the fairness of payment, and their flexibility. Further, we present results on the design feedback quality and quantity. In the third part, we present the themes that resulted from our qualitative interviews.

5.1 CrowdSurfer Usage Behavior

In this section, we describe the crowdworkers' behavioral interaction with the CrowdSurfer based on our log data. The results are displayed in Table 1. In total, participants solved 240 tasks of the CrowdSurfer within the experimental period of seven days. 15 of these tasks only contained a star rating. While 50 crowdworkers installed the CrowdSurfer, only 41 provided feedback at least one time. The majority of crowdworkers provided feedback between two and seven times.

The most frequently used feature after the submit button was the toggle button which turns the CrowdSurfer off. In this mode, no feedback pop-ups are displayed. Crowdworkers solved more than half of the tasks in the first two days after installing the CrowdSurfer (see Figure 5). Of the eight websites on which tasks were available, crowdworkers provided the most feedback on YouTube, followed by Amazon and Weather.com. Although crowdworkers did not complete every task the first time it was displayed, for 87.59% of the displayed tasks crowdworkers submitted feedback, eventually. On average, crowdworkers submitted tasks 69.23 seconds (SD = 49.20 seconds) after entering the website.

5.2 Working Conditions & Feedback Quality

5.2.1 Working Conditions of Crowdworkers. To assess the crowdworkers' perceptions, we analyzed the responses to questionnaire items. To assure the internal consistency of latent constructs, we assessed outer factor loadings and Cronbach's alpha with a cutoff at 0.7 and 0.6 [17, 45]. Afterward, scales were mean-scored. To assess

Table 1: CrowdSurfer feature usage by crowdworkers.

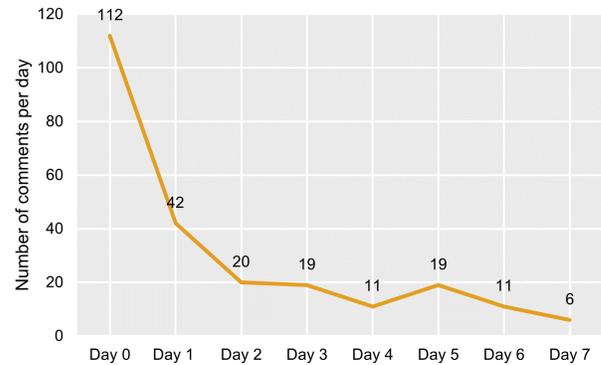| Feature	Used ... times	Used by ... crowdworkers	Average usage per worker
Task information	4	3	1.33
Reminder	7	3	2.33
Minimize	10	8	1.25
On/off task search	96	36	2.67
Feedback submit	240	41	5.85

Figure 5: Submitted tasks per day over the period of seven days.

the effect of the experimental treatment conditions (baseline vs. CrowdSurfer), we conducted an analysis of variance (ANOVA) for each variable and the feedback quality and quantity assessments as dependent variables. The results show no significant results for work flexibility, but a significant effect for fairness of payment ($F_{(1,61)} = 5.76$, $p < 0.05$) between the CrowdSurfer and the baseline treatment. For the perceived time invested to complete the task, we find a significant effect ($F_{(1,61)} = 4.02$, $p < 0.05$). Detailed information regarding descriptive statistics is presented in Table 2. To complement the quantitative analysis, we present boxplots of the perceptive measures in Figure 6.

5.2.2 Design Feedback Quality. For the six design feedback quality dimensions, we performed ANOVAs to test the effect of the treatment on the dependent variables for the feedback comments. For almost all variables we see a positive main effect between the baseline and the treatment group with higher values for the baseline condition (see Table 3 for detailed results of the ANOVA tests and Figure 7 for the boxplots). Only for positivity, we see a higher level in the CrowdSurfer condition and a not significant main effect ($F_{(1,563)} = 0.058$, $p = 0.81$). Further, we analyzed the difference in the length of the feedback comments provided by the participants. To do so, we analyzed the number of characters per comment. The results of the ANOVA showed a significant main effect ($F_{(1,563)} = 9.26$, $p = 0.01$) with longer comments in the baseline condition.

5.3 CrowdSurfer Experience

We analyzed and coded the interviews to understand how crowdworkers perceived the CrowdSurfer for conducting feedback tasks.

Table 2: Descriptive statistics of perceptive measures over the two treatment conditions.
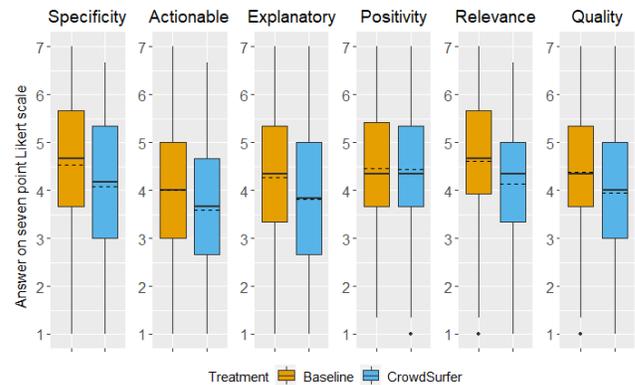
Dependent variable	Details	Baseline (n = 29)	CrowdSurfer (n = 34)	Analysis results
Work flexibility	Mean	5.335	5.273	Not significant, F(1,61) = 0.07, p = 7.87
	(SD)	(0.987)	(0.824)	
Fairness of payment	Mean	4.931	5.735	Significant, F(1,61) = 5.76, p < 0.05
	(SD)	(1.665)	(0.946)	
Perceived task	Mean	11.241	8.029	Significant, F(1,61) = 4.02, p < 0.05
completion time	(SD)	(6.098)	(6.530)	

Table 3: Statistics of design feedback quality dimensions over the two treatment conditions, aggregated on comment level.

Dependent variable	Details	Baseline (comment, n = 340)	CrowdSurfer (comment, n = 225)	Result
Specificity	Mean	4.529	4.078	Significant, F(1,563) = 14.16, p < 0.01
	(SD)	(1.367)	(1.441)	
Actionable	Mean	4.010	3.593	Significant, F(1,563) = 14.29, p < 0.01
	(SD)	(1.268)	(1.282)	
Explanatory	Mean	4.273	3.803	Significant, F(1,563) = 13.15, p < 0.01
	(SD)	(1.501)	(1.516)	
Positivity	Mean	4.455	4.430	Not significant, F(1,563) = 0.058, p = 0.81
	(SD)	(1.261)	(1.221)	
Relevance	Mean	4.609	4.133	Significant, F(1,563) = 19.61, p < 0.01
	(SD)	(1.239)	(1.273)	
Overall Quality	Mean	4.381	3.948	Significant, F(1,563) = 14.01, p < 0.01
	(SD)	(1.341)	(1.355)	
Comment Length	Mean	141.674	112.009	Significant, F(1,563) = 9.26, p < 0.01
	(SD)	(116.017)	(109.458)	
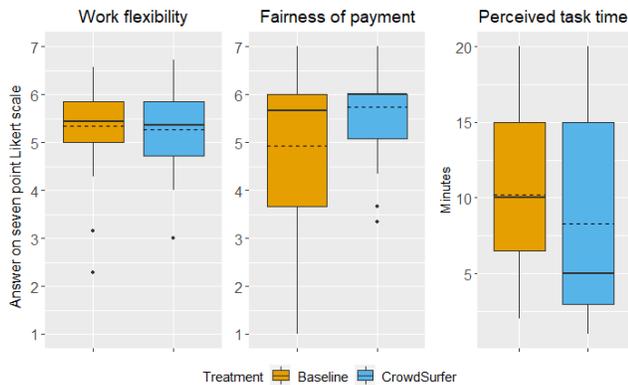
Figure 6: Boxplots of perceptions of work flexibility, fairness of payment, and time for task completion (the dotted line represents mean value).

Figure 7: Boxplots of design feedback quality dimensions (based on feedback comment level, dotted line represents mean value).

We derived 20 themes that describe crowdworkers' positive and negative experiences with the CrowdSurfer. Overall, all participants liked the concept of the CrowdSurfer. They found conducting tasks with the CrowdSurfer easy and fun, and experienced the interaction as comfortable. However, they still experienced issues and raised concerns, especially regarding the clarity of the tasks and the security of their personal data when installing an extension. An

overview of the positive and negative aspects of the CrowdSurfer based on our interviews is shown in Table 4.

5.3.1 CrowdSurfer Usability. In the first part of our interviews, we asked participants how they perceived the interaction with the CrowdSurfer in general. Ten participants mentioned that they appreciated that the CrowdSurfer was so easy to use and perceived the interaction as very comfortable (*"It was simple, I mean anyone*

Table 4: Overview of positive and negative aspects of the CrowdSurfer derived from the qualitative interviews.

	Positive aspects	Negative aspects
CrowdSurfer usability	- simple to use - fun to do tasks - comfortable and organized - the combination of quantitative and qualitative feedback - more personal feedback requests	- requests are not specific enough - UI design could be improved - security and data privacy concerns
Feedback process	- seamless integration in normal internet surfing - more in the usage context - transparent regarding what and when data is collected	- interruption of other tasks - not all tasks will be found - could get repetitive
Working conditions	- no pressure in terms of if, when and how to do tasks - payment higher than expected - less effort for background work - no feeling of scarcity - multiple incentives besides monetary reward	- some participants searched for tasks - no trust that feedback will be used

could do it. There's nothing technical about this. It just asks a question and you give your answer" (P4)). Three participants stated "it was fun" (P5) to provide feedback with the CrowdSurfer.

They perceived the CrowdSurfer as a non-intrusive, transparent, and seamless way to provide feedback ("I knew it was there in the background. That was one thing that was good. It wasn't hiding." (P4)). The feedback requests felt for three participants very personal. P5 explains "I think it also kind of feels a little more personalized in a way because it's not just like this survey form that you fill out and everybody fills out the same thing. Like when it pops up on your screen, while you're doing the browsing, it kind of feels more intimate [...]". On the negative side, five participants reported that they were worried about the security and their data privacy when installing the CrowdSurfer: "I think the only issue is when people might think 'Well, hang on a minute, can I trust this to be on all the time, or should I turn it off when I'm banking or checking personal information?'" (P3). However, overall they perceived the CrowdSurfer as trustful enough to still decide to install the extension. Four times participants mentioned that they would like to have more background information about why and by whom the feedback is collected to better target their feedback because "what was there was quite basic" (P4) and "it can be a little bit ambiguous because you could always be rating different facets of whatever X is" (P12). Furthermore, they perceived the feedback requests as "vague" (P6). While it was seen as positive that the feedback pop-ups blend in with the original website, two participants were worried that they would miss feedback requests because the pop-ups do not stand out enough. One participant felt like the UI design could be "a little more advanced" (P5).

CrowdSurfer vs. traditional crowdsourcing tasks. Seven participants explained that the main difference between providing design feedback via the CrowdSurfer and doing it in a survey is that they felt more in the context of use: "I think that I like this better than just filling out a normal survey because when they're asking questions it's about what I'm seeing right there in front of my eyes, so I don't have to rely on my memory of the experience [...]. I'm in the experience, I can read it, I can say what I think." (P5). The feedback situation is "more direct" (P2), and questions are asked "at the relevant time"

(P1). The crowdworkers are not "in the mindset of being paid to go through a website and break it down and try to find things wrong with it" (P11) and are therefore able to "give a more authentic answer" (P11). They even spent less time thinking about their feedback, which made them feel like their feedback gets more valuable. Three participants perceived the CrowdSurfer as being more comfortable than filling out a survey because although they thought the pop-ups were surprising and random, the tasks felt more predictable "in terms of when and how many tasks you might do" (P12). Additionally, answering the CrowdSurfer questions felt like less effort than doing the same in a survey.

CrowdSurfer features. Regarding the features, four participants in our interviews mentioned that they used the on-off toggle button. They used it either to turn off the CrowdSurfer when they did not want to be traced or interrupted or to refresh the available tasks. They thought when turning the CrowdSurfer on again new tasks might pop up ("I just wanted to see, if it's gonna be giving me tasks if I switch it on and off" (P6)). Five participants used the overview to check their rewards or last tasks. Two participants liked that they could see how many tasks they already found and solved on the current website so that they knew when they could stop looking out for tasks. P4 explained: "I saw that I'd done two out of two tasks, so I knew I didn't have to go around and browse on Amazon anymore. It was done and dusted. [...] You know exactly where you stood." The reminder and minimize functionalities were not used very frequently. Nevertheless, four participants still found that these might be useful for situations in which they "didn't have the time or [...] didn't have the mind to take a pause [...]" (P9).

Ideas for improvement. Addressing the perceptions that the feedback requests and the overview panel could have been more detailed, participants recommended features "to compare with other people who use it" (P4), "a little history of what was going on" (P4), or "an up to date list of all of the sites that you could sign in" (P2). Regarding the list of available tasks, P2 argued "you wouldn't have the problems of random sort of winning the lottery by getting a website where there is a question. If people were expecting to actually earn the money regularly doing this task, I think they'd have to have the structure of a list rather than the frustration of just sort of wandering around and

hoping that one of the sites was on the list [...] I would see that as a waste of my time, and I'm not sure I would take it seriously. I think I'd go and do something else". Four participants also asked for more interactive ways to provide feedback ("Either use phrases that people can choose from or numbers, or they can drag their mouse from one point to the other just kind of engage people in different ways, you can more interact. If people prefer one type of feedback over the other then at least have that variety." (P5)). As two participants stated that they sometimes accidentally submitted their feedback too early they asked for a way to call back the feedback. P1 stated that it "would be useful to have that as a feature where you can go 'hang on, I forgot to say this'". Finally, P4 suggested making the CrowdSurfer more intelligent so that it recognizes when the user is willing to provide feedback.

5.3.2 Feedback Process. Mainly there are two different users types: Either, they want to solve the tasks as quickly as possible and actively searched for the tasks ("I don't think I would simply wait for something to randomly appear. If I've promised to do a task I like to have a list of what the expectations are and go and do them" (P2)), or they waited for tasks to pop up during their everyday internet surfing ("I didn't get to the point where I had to search for a task" (P7)). P3 stated: "I didn't modify my behavior because the crowdsurfing app was there. I just did my normal thing". The group of participants who waited for tasks to pop up was much larger than the other one. Five participants declared that they provided feedback usually directly after they saw the pop-up: "It popped up and just straight away I put in the information" (P4). Thereby, they mainly shared their quick and immediate reaction to the question because they believed "often a quick response is the right one" (P3). Four participants stated that they always provided feedback when they noticed a feedback request. For example, P6 disclosed: "I didn't decide. I just had to do it for each task that I was given. Like there's none that I saw, and I was like 'No, I'm not doing this one.'". Situations in which they did not provide feedback were when they were "really in a hurry" (P3) or on websites, they "consider to be unpleasant" (P2). One drawback for five participants was that they could get interrupted by feedback requests when doing important primary tasks. P11 explained her concerns as follows: "I could see it get a little bit frustrating because I'm here on Amazon because I need to buy something and Amazon is distracting enough to have another thing pop up and inhibit my shopping process".

5.3.3 Working Conditions.

Fairness of Payment. Although multiple participants mentioned that they actively searched for tasks, four participants felt the invisible work to be less than in traditional crowdsourcing tasks ("With this one, it was easier because [...] the only thing I had to do was to review. The [demographic] background information is already there." (P6)). Three participants experienced that solving tasks with the CrowdSurfer requires less effort and time for preparation before the actual task because "there's less background work that needs to be done" (P1). One participant was even surprised about how much money she made when checking the rewards for the first time.

Flexibility. Overall, they perceived the CrowdSurfer setup as very flexible. Six participants liked that it felt not pressured ("It was super chilled. There was no pressure in terms of time and I could do

it whenever I wanted [...]. So it was super comfortable, better than the Prolific site" (P6)) or if to work at all as "there is no penalty for not giving feedback" (P11). In contrast to doing tasks on Prolific, they did not have to be online at a specific time when new tasks are published. This reduced "that feeling of scarcity around it" (P11). They liked that "[...] there is a steady supply of work that could be done" (P12). However, some workers also stated that the task setup did not have an impact on their flexibility in doing tasks. They also did not see a significant impact on their work-life balance, as P3 explained: "I literally just did my normal day, nothing to do with work/personal life balance, nothing like that was affected by it at all". This was mainly because they could turn it off when they did not want to be interrupted: "I think I would like both options to be available to me and that I'd be able to choose, and for that choice to be inconsistent. So like one day if I feel like I want to browse and I want to also be able to make some money on the side, then I'd be able to toggle it on to activate it [...] Some other days I might feel like [...] I don't want anyone to be asking me things [...], so I'd be able to have it off, and then it wouldn't pop up. But I think both options can be very useful." (P5).

Motivation. The main reason for providing feedback was the monetary reward ("Mostly it was for money" (P6)) as mentioned by seven participants. However, participants also liked that they were able to share their opinion (four participants), help us with our study (three participants), be able to improve the websites (six participants), or were just curious (two participants). One participant also liked that she now "actually understood what it takes to write a review" (P6). Participants felt like they could make an impact with their feedback by contributing to a bigger picture. However, some workers did not care about the impact. Although they felt quite competent to provide meaningful design feedback, especially for websites they visit frequently, two participants mentioned that they would be able to provide better feedback if more background information on the task was provided. They had questions like "What is she specifically looking for here? [...] What will he use the feedback for? Why is it important to be concerned about the colors?" (P6). Also, they felt that the feedback pop-up did not encourage them to be reflective, as P3 phrased: "It didn't encourage me to be reflective. It kind of encouraged me to give a quick response". It helped three participants that they already had an opinion for the websites that they were familiar and they "just answered the question based on [their] experiences" (P3). Additionally, the tasks were so easy that everyone could do them. Detrimental for the motivation of two participants was that they "don't really trust companies that ask for feedback in general because they never act upon it" (P3).

6 DISCUSSION

The majority of the crowdworkers using the CrowdSurfer perceived the provision of feedback as more comfortable, simple, fun, and personal than in a traditional design feedback survey. Further, these crowdworkers perceived the payment as fairer and spent less time on the task. On the other side, the feedback collected with the CrowdSurfer was less specific, actionable, and relevant, contained fewer explanations, and was of lower quality. Our participants mentioned potential reasons for the reduced feedback quality such as the divergence between a primary and a secondary task in the

CrowdSurfer treatment. In the following, we discuss three essential theoretical and practical implications of our study and present design recommendations for the design of crowd-feedback systems for everyday internet surfing.

6.1 Integrating Crowdsourcing Tasks in Crowdworkers' Everyday Internet Surfing Leads to Less Effort

Over the years, many researchers have argued for higher payments of crowdworkers [18], especially considering the balance of effort and payment [26]. Further, they advocate for more flexible working conditions [30, 48]. In our study, one main effect of the CrowdSurfer was its positive impact on these working conditions such as the fairness of payment, the time spent solving tasks, and work flexibility. Participants stated in the interviews that the seamless integration made it easier and created less effort for them to provide feedback compared to traditional tasks on crowdsourcing platforms. Our quantitative survey results reveal that crowdworkers actually spent less time working on feedback tasks and perceived the payment as fairer than those in the baseline treatment. These two results make sense. Perceiving tasks to be of less effort makes the payment seem higher and thereby fairer. However, our results are twofold regarding the work effort. Although the time for searching for tasks was significantly lower than in the baseline treatment and most participants reported that they had no invisible work, a few workers reported that they actively searched for tasks. However, searching for tasks was not possible in the baseline. Thinking about a long-term scenario in which the CrowdSurfer continuously offers new tasks to crowdworkers, searching for tasks would become even more counterproductive, especially when the tasks are published on less popular websites.

Crowdworkers did not feel like the CrowdSurfer had a negative impact on their flexibility or their work-life balance. The main reason for this was the functionality to turn it on or off whenever they like. The on/off feature did not only help them to ensure that they are not interrupted when working or doing other important tasks but also allowed them to guarantee that their interaction and data are not tracked when surfing privately on the internet (cf. for internet banking). Consequently, we believe the on/off feature is a core element that made crowdworkers feel flexible and comfortable when working with the CrowdSurfer. To sum it up, the CrowdSurfer not only leads to less effort for crowdworkers but also offers a higher hourly wage to crowdworkers and allows them to be more flexible when working on tasks.

6.2 The Quality of In Situ Feedback is Lower than in Dedicated Surveys but the Feedback is More Real

Our data shows that the feedback quality is in most dimensions worse when collecting feedback via the CrowdSurfer. According to our interviews, crowdworkers believe that their feedback is still more valuable and real when they provide it in situ and for websites they frequently use. But why is the feedback quality worse? Why does the real usage scenario not lead to more relevant and actionable

feedback? Is our proposed approach still a successful model for crowdsourcing design feedback?

First, a potential explanation for the reduced feedback quality is the shorter length of the feedback comments that contain fewer details. One reason for shorter comments could be the different sizes of the text fields for feedback comments. In the baseline, the text field was bigger, which might have led crowdworkers to think they need to write more. Second, in our interviews, participants stated that they mainly shared their quick reactions to the question. This is consistent with the log data that showed that on average the feedback tasks were submitted about one minute after the participants entered the website. They unconsciously provided quick feedback which came directly "from the heart". The main difference between the two treatments was that in the baseline providing feedback was the primary task on which participants were focused, while when using the CrowdSurfer providing feedback was a secondary task, and participants potentially focused on another primary task. We assume that the CrowdSurfer treatment group spent overall less time and effort on the feedback provision process as the initial effort was close to zero. Compared to the baseline treatment, they did not have to spend time entering a website, getting familiar with the element on which the feedback is collected, and forming an opinion. They only had to document their thoughts and perceptions. As we learned in our interviews, participants were not only motivated by monetary compensation. Consequently, they might care less about receiving the monetary reward and, in turn, put less focus on their feedback quality, and more on the feedback honesty.

Third, one potential side-effect of the CrowdSurfer might be that it favors a special character of crowdworkers. Crowdworkers who are willing to install an extension and are open to an innovative task form might have special approaches to crowd work. They are more flexible and might be less focused on maximizing their financial outcomes. We assume that these workers do not use Prolific as their primary source of income, but are rather "part-time" crowdworkers.

Finally, our interviewees mentioned that CrowdSurfer users did not feel like being paid for finding problems on a website like in the baseline survey. The focus on the problem-finding task itself might have created the perceived urgency to report design flaws. This could be explained by a social desirability bias [15]. Participants know that they are explicitly recruited for highlighting design errors. In consequence, they come up with issues, even though these might not represent actual impressions. In the CrowdSurfer, the collected feedback on issues is more subconscious and, thus, closer to participants' perceptions. One might say the feedback is more real and less biased and therefore superior to the traditional survey design feedback. Drawing on Goncalves et al. [13], exploring motivational factors besides money could be a useful approach to increasing the feedback quality and understanding what makes crowdworkers report design issues besides the monetary reward. Our results also align with related studies on integrating secondary crowdsourcing tasks in primary tasks [16]. Although we followed a different motivation, we also come to the conclusion that integrating feedback tasks into crowdworkers everyday surfing is overall a successful way to accomplish meaningful design feedback. Consequently, we argue that the CrowdSurfer is a valuable approach to collecting honest and unbiased design feedback in comparison to traditional surveys.

6.3 Crowdworker Archetypes: Seamless Integration vs. Waiting for Tasks

The CrowdSurfer was intended and designed for crowdworkers who want to solve tasks and earn money while doing other primary tasks and therefore, we selected Prolific as a platform that the majority of workers are not using as a primary source of income [1]. The results of our interviews indicate that there exist two types of crowdworkers: Either the crowdworkers liked being able to solve tasks during their everyday internet browsing and did not adjust their browsing behavior because of the CrowdSurfer, or they did not like the random appearance of tasks popping up on websites. The second group of crowdworkers actively searched for tasks. They identified the relevant websites by searching in the experimental descriptions where we stated that a requirement for participation is that they frequently visit some of the mentioned websites. These two groups of workers can also be linked to the work–nonwork boundary management profiles of humans [29]. There are humans who like to integrate work tasks and non-work tasks, while there are also workers who find it difficult to set appropriate boundaries to not get interrupted. Currently, the CrowdSurfer design mainly serves the so-called fusion lovers [29]. They liked to surf the internet and earn money during this activity. The second group of crowdworkers still liked to execute tasks at hand. Although these participants did not have the real usage scenario as it was intended for the CrowdSurfer, they still saw advantages in the browser extension. We assume, that this group preferred to separate work and non-work tasks and respectively set their boundaries. Similar to the first group, they also felt more in the context when providing feedback and thought that it is more seamless and less effort to use the extension to provide feedback than doing it via a survey. However, the additional search process for tasks might have confused them and increased their invisible work.

Therefore, we think it would be desirable to address both types of workers in the future. To do this, the CrowdSurfer could offer a list of available tasks. This also has the advantage for feedback requesters that crowdworkers could be guided to new or less frequently visited websites to provide feedback.

6.4 Design Recommendations for Browser Extensions to Integrate Crowdsourcing Tasks in Everyday Internet Surfing

Based on the results and the implications that we discussed in the sections above, we derived six design recommendations for the design of browser extensions to integrate tasks into crowdworkers' everyday internet surfing. These insights shall help future researchers to design similar extensions for other types of tasks. Following the structure that we used to analyze the qualitative interviews, each recommendation is assigned to one of the three concepts: *Usability, Work Process,* and *Working Conditions* of crowdworkers.

(1) *Present users an overview of the collected data (Usability).* Showing users which data is collected about them increases the transparency of the extension which in turn positively affects users' trust. Further, users can better manage their tasks and rewards.

(2) *Provide support, guidance, and background information (Usability).* Participants in our study stated that they believe that they would have been able to provide even better feedback if they had more background information on the tasks or support in solving the tasks. The extension should provide users with important information about the requirements of a task and provide them support in solving the tasks.

(3) *Ensure task conduction is quick and easy to limit interruption of users in daily life (Work Process).* The integration of crowdsourcing tasks only makes sense when the tasks are simple and quick to complete. When users have to spend more time than a few minutes to solve the tasks, they might feel interrupted in their actual task and refuse to do it.

(4) *Offer a way to actively search for tasks (Work Process).* We learned that there exist crowdworkers who do not like waiting for tasks to pop up and prefer a list of available tasks to complete. There should be an option for these crowdworkers to actively search for and directly access available tasks whenever they are willing to work.

(5) *Support on/off functionality for the browser extension (Working Conditions).* Offering users to turn off the extension when they exclusively want to surf privately on the internet is important. Users need to keep their flexibility between work and private time. Also, it helps to increase trust in the extension as crowdworkers can turn it off when they do not want to be traced.

(6) *Make all tasks voluntary and allow the rejection of tasks (Working Conditions).* Workers liked in our study that they could freely decide which tasks they want to do and were not forced to do tasks on websites they did not feel comfortable doing. Forcing workers to do tasks might reduce their willingness to participate in the tasks at all.

7 LIMITATIONS & FUTURE WORK

In this section, we summarize the limitations that we acknowledge in this study and connect them to future research avenues. Further, we present our vision of how feedback requesters and crowdworkers could use the CrowdSurfer in practice.

First of all, we caution against overgeneralizing the findings from this CrowdSurfer study. Our findings are limited by the self-selecting sample of participants caused by our study design and by the websites and feedback tasks that we selected. Also, our study did not present a real feedback scenario and participants could experience the CrowdSurfer only for seven days and not in the long term. Due to the nature of the experiment and the innovativeness of the CrowdSurfer as a browser extension, this was not the case for our study. However, we believe that our results already provide good indicators for the applicability to continuously collect feedback. Overcoming this limitation requires longer user studies. While we believe that new feedback tasks need to come from real feedback requesters future work needs to bring the CrowdSurfer to life, connect it to real feedback requesters, and investigate its effects in the wild.

Second, as mentioned in the discussion, some participants actively searched for tasks instead of waiting for them. They were in a working mode and did not want to wait for tasks. The CrowdSurfer

did not provide functionalities for these users to directly access the tasks at hand. Future work should derive two actions: First, (1) investigate how the feedback differs between crowdworkers who actively searched for tasks and crowdworkers who did not. Second, the CrowdSurfer should be designed to (2) allow users to find tasks easily and simplify the search for tasks. To do so, the CrowdSurfer could provide a list of all available tasks. Further, this list would also simplify the feedback collection for less frequently visited websites.

Third, one core element of our study was the assessment of feedback quality. We recruited crowdworkers with UI design experience to evaluate the collected feedback. Of course, this might lead to different results than asking the actual website designers and developers how valuable they perceive the feedback. Still, this method was already applied in similar studies, e.g., [20] and as the feedback in both treatments was evaluated in the same way, we believe that the results are overall still valid and comparable. In further studies, the feedback could be analyzed with the help of actual feedback requesters. We assume that these feedback requesters might have different expectations regarding the feedback than the quality dimensions that we analyzed. Further, a clear explanation of feedback requesters' requirements and the benefits of real in situ user feedback compared to traditional survey-based feedback would help to design further crowd-feedback systems.

Finally, we designed the CrowdSurfer to conduct tasks for design feedback. However, on crowdsourcing platforms, there are multiple other types of tasks available (e.g., matching, labeling, idea creation, captioning). For being a potential tool to conduct crowdwork, the CrowdSurfer should allow further task types. In particular, task types that rely on internet usage and context seem prone to the CrowdSurfer application. Future research should investigate which task types are applicable to present with the CrowdSurfer and beneficial for both the task requesters and the crowdworkers.

7.1 CrowdSurfer Implementation Concept

While our CrowdSurfer field study aimed to replicate a realistic feedback scenario, there are several elements to be considered when bringing the CrowdSurfer into the real world beyond an experimental use case. In this section, we describe a potential implementation concept for the usage of the CrowdSurfer in a real-world setting for the crowdsourcing of design feedback.

Overall process. Figure 8 presents the structure and usage flows for feedback requesters and providers (crowdworkers). There are two main intertwined processes within the CrowdSurfer application, one for the feedback requesters and one for the feedback providers. First, feedback tasks need to be created by requesters, and, second, they need to be fulfilled by crowdworkers. For the first process, feedback requesters visit the CrowdSurfer website (A) and can, in very simple user interaction, create the task by providing the URL of the respective website, the HTML ID attribute of the element on which feedback is required, and a question that feedback providers shall answer. They can, further, indicate how much feedback they want. Feedback requesters also pay via this requester interface for their feedback requests. The payment includes the rewards for the feedback providers, the service fees for the crowdsourcing platforms, and a service fee for the CrowdSurfer operators. The requested task and respective payment are then stored in the CrowdSurfer database (1). At the same time, CrowdSurfer tasks are continuously presented on crowdsourcing platforms (e.g., Amazon Mechanical Turk, Prolific) (B). These tasks allow the CrowdSurfer operator to recruit new crowdworkers, whenever they are needed. In these CrowdSurfer tasks, crowdworkers are asked to install the CrowdSurfer extension and log in with their respective crowdsourcing platform ID (2). In return, they receive a fixed payment and the task is completed. Now, whenever a CrowdSurfer user visits a website on which tasks are available (3), they see the feedback pop-up (C). When they submit feedback, their answer is stored in
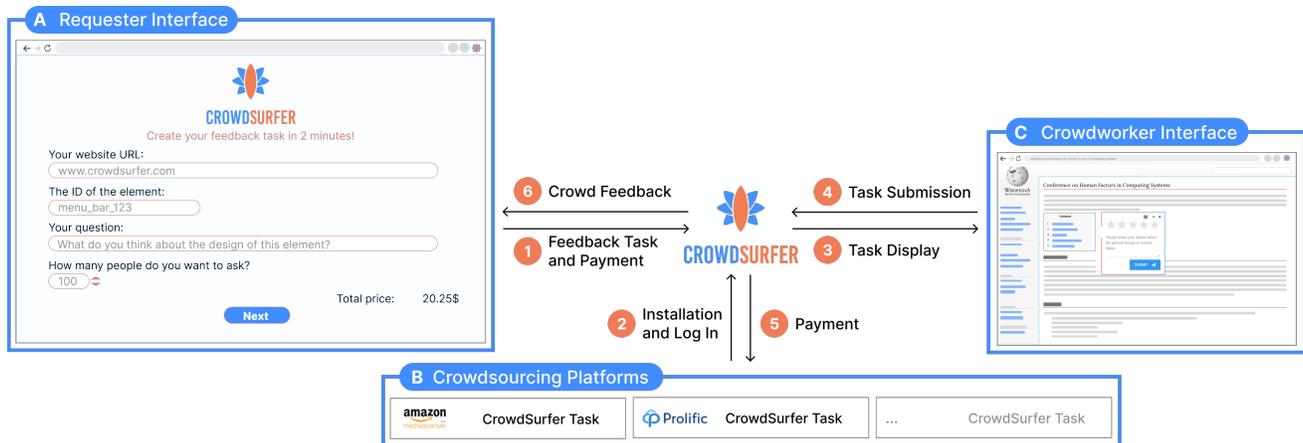

Figure 8: The CrowdSurfer process as we envision it for implementation in practice. 1) First, requesters create tasks that are added to the CrowdSurfer database. 2) Crowdworkers are continuously recruited to install and set up the extension. 3) Available tasks are then displayed on the websites whenever a crowdworker accesses them. 4) The submitted task is stored in the CrowdSurfer database. 5) The crowdworkers are paid via weekly bonuses using the initial installation tasks. 6) The received feedback is published to the requester interface.

the CrowdSurfer database (4). As the mentioned crowdsourcing platforms allow the payment of bonuses, the CrowdSurfer operator will use this functionality to pay CrowdSurfer users for their feedback on a weekly basis (5). Feedback requesters can then access the submitted feedback via the feedback requester interface on the CrowdSurfer website (6).

Task assignment. The task assignment will be based on a first-come-first-serve functionality. That means when a new task is submitted with 100 feedbacks requested, the first 100 workers who access this page will see the feedback pop-up. When workers minimize a task, the task is reserved for them as long as they stay on this website. Workers can also reserve a task by setting a reminder one time for 24 hours. When they reject a task, the task is passed on to the next CrowdSurfer user who visits the website.

CrowdSurfer for less popular websites. We are aware that the proposed concept requires a sufficiently large and heterogenous CrowdSurfer user base to work. We also acknowledge that less popular websites will not be challenged to collect a meaningful amount of feedback via this approach. A possible solution to address this problem is to notify crowdworkers whenever a task is available on a website that is similar to the one that they are currently on. This can be done via the CrowdSurfer panel and for example be integrated into the task list, that we mentioned in the ideas for further improvements of the CrowdSurfer. To do this, feedback requesters must choose a website category to which their website belongs when creating a feedback request. Thereby, the CrowdSurfer favors tasks on websites that are less frequently visited and still lack a larger number of feedback submissions. This allows us to make sure that also less frequently visited websites will receive feedback submissions while at the same time we can ensure that the context is always given when providing feedback. So for example, whenever the crowdworker visits a shopping website, available tasks on other shopping websites are added to the list in the CrowdSurfer overview panel. When these tasks are completed by the required number of crowdworkers, the crowdworker has completed this task himself, or the last visit to a shopping website was more than an hour ago, the task is deleted from the list.

8 CONCLUSION

Real user feedback is a valuable means to evaluate and continuously improve website designs. However, it often does not lead to the desired amount of feedback. Crowdsourcing of design feedback is a scalable alternative but also comes with drawbacks. In our study, we aimed to provide a seamless approach to crowdsource in situ design feedback. Besides developing design rationales and recommendations for the integration of crowdsourcing tasks into crowdworkers' everyday internet surfing, we wanted to understand how the in situ feedback from real users differs from traditional crowd feedback on Prolific and how crowdworkers perceive the innovative approach for conducting crowdsourcing tasks. Therefore, we developed the CrowdSurfer, an innovative crowd-feedback system as a Chrome browser extension, based on exploratory interviews with crowdworkers. To analyze these effects, we conducted a field study over seven days in which crowdworkers could use the CrowdSurfer to provide design feedback on eight popular websites.

We compared the resulting design feedback and quantitative answers of a post-task questionnaire with a traditional survey-based feedback collection. Further, we analyzed the resulting feedback and conducted twelve semi-structured interviews to understand the CrowdSurfer experience from a crowdworker perspective. Our results show that crowdworkers enjoyed our innovative Crowd-Surfer design, felt more in the experience, perceived the effort to be lower than in a survey, and expected their feedback to be more relevant. Nevertheless, the feedback quality was lower. Our findings demonstrate the feasibility of integrating tasks into crowdworkers everyday internet surfing. Still, they show that offering a more effortless way to provide feedback in return for a monetary reward might also have a negative impact on feedback quality. Our results motivate further investigations for the design of similar crowdsourcing tasks. Overall, we contribute with our work to enhance the feedback collection processes while improving the working conditions for crowdworkers.

REFERENCES

[1] Tahir Abbas and Ujwal Gadiraju. 2022. Goal-Setting Behavior of Workers on Crowdsourcing Platforms: An Exploratory Study on MTurk and Prolific. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 10, 1 (10 2022), 2–13. https://doi.org/10.1609/HCOMP.V10I1.21983

[2] M Almaliki, C Ncube, and R Ali. 2014. The Design of Adaptive Acquisition of Users Feedback: An Empirical Study. In *2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*. IEEE, New York, NY, USA, 1–12. https://doi.org/10.1109/RCIS.2014.6861076

[3] Paul Alpar and Lars Osterbrink. 2018. Antecedents of Perceived Fairness in Pay for Microtask Crowdwork. In *26th European Conference on Information Systems, ECIS 2018*. Association for Information Systems (AIS), New York, NY, USA, 1–13.

[4] Oshrat Ayalon and Eran Toch. 2018. Crowdsourcing Privacy Design Critique: An Empirical Evaluation of Framing Effects. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, Vol. 2018-Janua. Association for Computing Machinery (ACM), New York, NY, USA, 4752–4761. https://doi.org/10.24251/hicss.2018.598

[5] Oshrat Ayalon and Eran Toch. 2019. A/P(rivacy) Testing: Assessing Applications for Social and Institutional Privacy. In *Extended Abstracts of the 2019 CHI Conference*. Association for Computing Machinery (ACM), New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312972

[6] Janine Berg. 2016. Income security in the On-Demand Economy: Findings and Policy Lessons from a Survey of Crowdworkers. *Comparative Labor Law and Policy Journal* 37, 3 (2016), 27.

[7] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[8] Wen Chih Chiou, Chin Chao Lin, and Chyuan Perng. 2010. A Strategic Framework for Website Evaluation based on a Review of the Literature from 1995-2006. *Information and Management* 47, 5-6 (8 2010), 282–290. https://doi.org/10.1016/j.im.2010.06.002

[9] Cristiano Codagnone, Fabienne Abadie, and Federico Biagi. 2017. The Future of Work in the Sharing Economyy. Market Efficiency and Equitable Opportunities or Unfair Precarisation? *SSRN Electronic Journal* JRC101280 (2017), 96. https://doi.org/10.2139/ssrn.2784774

[10] James R. Detert, Linda Klebe Treviño, and Vicki L. Sweitzer. 2008. Moral Disengagement in Ethical Decision Making: A Study of Antecedents and Outcomes. *Journal of Applied Psychology* 93, 2 (3 2008), 374–391. https://doi.org/10.1037/0021-9010.93.2.374

[11] Alek Felstiner. 2011. Working the Crowd: Employment and Labor Law in the Crowdsourcing Industry. *Berkeley Journal of Employment and Labor Law* 32, 1 (8 2011), 19–24. https://doi.org/10.3109/9780203490891-6

[12] Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2007. MyExperience: A system for In Situ Tracing and Capturing of User Feedback on Mobile Phones. In *MobiSys'07: Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*. Association for Computing Machinery (ACM), New York, NY, USA, 57–70. https://doi.org/10.1145/1247660.1247670

[13] Jorge Goncalves, Denzil Ferreira, Simo Hosio, Yong Liu, Jakob Rogstadius, Hannu Kukka, and Vassilis Kostakos. 2013. Crowdsourcing on the Spot: Altruistic Use of Public Displays, Feasibility, Performance, and Behaviours. In *UbiComp 2013 - Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery (ACM), New York,

NY, USA, 753–762. https://doi.org/10.1145/2493432.2493481

[14] Jorge Goncalves, Simo Hosio, Jakob Rogstadius, Evangelos Karapanos, and Vassilis Kostakos. 2015. Motivating participation and improving quality of contribution in ubiquitous crowdsourcing. *Computer Networks* 90 (2015), 34–48. https://doi.org/10.1016/j.comnet.2015.07.002

[15] Pamela Grimm. 2010. Social Desirability Bias. In *Wiley International Encyclopedia of Marketing*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444316568.wiem02057

[16] Nathan Hahn, Shamsi T. Iqbal, and Jaime Teevan. 2019. Casual Microtasking: Embedding Microtasks in Facebook. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3290605.3300249

[17] Joe F. Hair, Marko Sarstedt, Lucas Hopkins, and Volker G. Kuppelwieser. 2014. Partial Least Squares Structural Equation Modeling (PLS-SEM): An Emerging Tool in Business Research. , 106–121 pages. https://doi.org/10.1108/EBR-10-2013-0128

[18] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. 2018. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574

[19] Kotaro Hara, Kristy Milland, Benjamin V. Hanrahan, Chris Callison-Burch, Abigail Adams, Saiph Savage, and Jeffrey P. Bigham. 2019. Worker Demographics and Earnings on Amazon Mechanical Turk: An Exploratory Analysis. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312970

[20] Saskia Haug, Ivo Benke, and Alexander Maedche. 2023. Aligning Crowdworker Perspectives and Feedback Outcomes in Crowd-Feedback System Design. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW (2023), 1–28. https://doi.org/10.1145/3579456

[21] Saskia Haug and Alexander Maedche. 2021. Crowd-Feedback in Information Systems Development: A State-of-the-Art Review. In *Proceedings of the 42nd International Conference on Information Systems (ICIS) 2021*. Association for Information Systems (AIS), New York, NY, USA, 17. https://aisel.aisnet.org/icis2021/is_design/is_design/4

[22] Saskia Haug and Alexander Maedche. 2021. Feeasy: An Interactive Crowd-Feedback System. In *Adjunct Publication of the 34th Annual ACM Symposium on User Interface Software and Technology, UIST 2021*. Association for Computing Machinery (ACM), New York, NY, USA, 41–43. https://doi.org/10.1145/3474349.3480224

[23] Danula Hettiachchi, Senuri Wijenayake, Simo Hosio, Vassilis Kostakos, and Jorge Goncalves. 2020. How Context Influences Cross-Device Task Acceptance in Crowd Work. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 8 (10 2020), 53–62. https://doi.org/10.1609/hcomp.v8i1.7463

[24] Lilly C Irani and M. Six Silberman. 2013. Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association of Computing Machinery (ACM), New York, NY, USA, 611–620. https://doi.org/10.1145/2470654.2470742

[25] Blake Ives and Margrethe H. Olson. 1984. User Involvement and MIS Success: A Review of Research. *http://dx.doi.org/10.1287/mnsc.30.5.586* 30, 5 (5 1984), 586–603. https://doi.org/10.1287/MNSC.30.5.586

[26] Toni Kaplan, Susumu Saito, Kotaro Hara, and Jeffrey Bigham. 2018. Striving to Earn More: A Survey of Work Strategies and Tool Use Among Crowd Workers. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. AAAI, 70–78. https://doi.org/10.1609/hcomp.v6i1.13327

[27] Anna Kasunic, Chun Wei Chiang, Geoff Kaufman, and Saiph Savage. 2019. Turker Tales: Integrating Tangential Play into Crowd Work. In *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 21–34. https://doi.org/10.1145/3322276.3322359

[28] Mehmet Kokoç. 2019. Flexibility in e-Learning: Modelling Its Relation to Behavioural Engagement and Academic Performance. , 16 pages.

[29] Ellen Ernst Kossek, Marian N Ruderman, Phillip W Braddy, and Kelly M Hannum. 2012. Work–nonwork boundary management profiles: A person-centered approach. *Journal of Vocational Behavior* 81, 1 (2012), 112–128. https://doi.org/10.1016/j.jvb.2012.04.003

[30] Laura Lascău, Sandy J.J. J Gould, Duncan P. Brumby, Anna L. 2022 Cox, Anna L. 2022 Cox, Laura Lascau, Sandy J.J. J Gould, Duncan P. Brumby, and Anna L. 2022 Cox. 2022. Crowdworkers' Temporal Flexibility is Being Traded for the Convenience of Requesters Through 19 'Invisible Mechanisms' Employed by Crowdworking Platforms. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/3491101.3519629

[31] Kurt Luther, Jari Lee Tolentino, Wei Wu, Amy Pavel, Brian P Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. Association for Computing Machinery (ACM), New York, NY, USA, 473–485. https://doi.org/10.1145/2675133.2675283

[32] Walid Maalej and Dennis Pagano. 2011. On the Socialness of Software. In *Proceedings - IEEE 9th International Conference on Dependable, Autonomic and Secure Computing, DASC 2011*. IEEE, New York, NY, USA, 864–871. https://doi.org/10.1109/DASC.2011.146

[33] James D McKeen and Tor Guimaraes. 1997. Successful Strategies for User Participation in Systems Development. *Journal of Management Information Systems* 14, 2 (1997), 133–150. http://www.jstor.org/stable/40398269

[34] Jonas Oppenlaender, Elina Kuosmanen, Andrés Lucero, and Simo Hosio. 2021. Hardhats and Bungaloos: Comparing Crowdsourced Design Feedback with Peer Design Feedback in the Classroom. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (ACM)*, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445380

[35] Jonas Oppenlaender, Thanassis Tiropanis, and Simo Hosio. 2020. CrowdUI: Supporting Web Design with the Crowd. *Proceedings of the ACM on Human-Computer Interaction* 4, EICS (2020), 1–28. https://doi.org/10.1145/3394978

[36] Dennis. Pagano. 2013. *PORTNEUF - A Framework for Continuous User Involvement*. Ph. D. Dissertation. Technical University Munich.

[37] Amy L. Richman, Janet T. Civian, Laurie L. Shannon, E. Jeffrey Hill, and Robert T. Brennan. 2008. The Relationship of Perceived Flexibility, Supportive Work-Life Policies, and Use of Formal Flexible Arrangements and Occasional Flexibility to Employee Engagement and Expected Retention. *Community, Work and Family* 11, 2 (5 2008), 183–197. https://doi.org/10.1080/13668800802050350

[38] Susumu Saito, Teppei Nakano, Chun Wei Chiang, Tetsunori Kobayashi, Saiph Savage, and Jeffrey P. Bigham. 2019. TurkScanner: Predicting the Hourly Wage of Microtasks. In *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*. Association for Computing Machinery (ACM), New York, NY, USA, 3187–3193. https://doi.org/10.1145/3308558.3313716

[39] Hanna Schneider, Katharina Frison, Julie Wagner, and Andras Butz. 2016. CrowdUX: A Case for Using Widespread and Lightweight Tools in the Quest for UX. In *DIS 2016 - Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. Association for Computing Machinery (ACM), New York, NY, USA, 415–425. https://doi.org/10.1145/2901790.2901814

[40] Thimo Schulze, Martin Schader, and Simone Krug. 2012. Workers' Task Choice in Crowdsourcing and Human Computation Markets. In *International Conference on Information Systems, ICIS 2012*, Vol. 4. Association for Information Systems (AIS), New York, NY, USA, 3551–3561.

[41] Norbert Seyff, Florian Graf, and Neil Maiden. 2010. Using Mobile RE Tools to Give End-Users their Own Voice. In *Proceedings of the 2010 18th IEEE International Requirements Engineering Conference, RE2010*. IEEE, New York, NY, USA, 37–46. https://doi.org/10.1109/RE.2010.15

[42] Norbert Seyff, Gregor Ollmann, and Manfred Bortenschlager. 2014. AppEcho: A User-Driven, In Situ Feedback Approach for Mobile Platforms and Applications. In *Proceedings of the 1st International Conference on Mobile Software Engineering and Systems*. Association for Computing Machinery (ACM), New York, NY, USA, 99–108. https://doi.org/10.1145/2593902.2593927

[43] Nada Sherief, Nan Jiang, Mahmood Hosseini, Keith Phalp, and Raian Ali. 2014. Crowdsourcing Software Evaluation. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. Association for Computing Machinery (ACM), New York, NY, USA, 1–4. https://doi.org/10.1145/2601248.2601300

[44] Melanie Stade, Marc Oriol, Oscar Cabrera, Farnaz Fotrousi, Ronnie Schaniel, Norbert Seyff, and Oleg Schmidt. 2017. Providing a User Forum is not Enough: First Experiences of a Software Company with CrowdRE. In *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference Workshops, REW 2017*. IEEE, New York, NY, USA, 164–169. https://doi.org/10.1109/REW.2017.21

[45] Ralf A.L.F. van Griethuijsen, Michiel W. van Eijck, Helen Haste, Perry J. den Brok, Nigel C. Skinner, Nasser Mansour, Ayse Savran Gencer, and Saouma BouJaoude. 2015. Global Patterns in Students' Views of Science and Interest in Science. *Research in Science Education* 45, 4 (8 2015), 581–603. https://doi.org/10.1007/s11165-014-9438-6

[46] Karel Vredenburg, Ji Ye Mao, Paul W. Smith, and Tom Carey. 2002. A Survey of User-Centered Design Practice. In *Proceedings of the 2002 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery (ACM), New York, NY, USA, 471–478. https://doi.org/10.1145/503376.503460

[47] Helen Wauck, Yu-Chun Yen, Wai-Tat Fu, Elizabeth Gerber, Steven P Dow, and Brian P Bailey. 2017. From in the Class or in the Wild?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery (ACM), New York, NY, USA, 5580–5591. https://doi.org/10.1145/3025453.3025477

[48] Mark E. Whiting, Grant Hugh, and Michael S. Bernstein. 2019. Fair Work: Crowd Work Minimum Wage with One Line of Code. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7 (10 2019), 197–206. https://ojs.aaai.org/index.php/HCOMP/article/view/5283

[49] Alex C. Williams, Gloria Mark, Kristy Milland, Edward Lank, and Edith Law. 2019. The Perpetual Work Life of Crowdworkers: How Tooling Practices Increase Fragmentation in Crowdwork. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (11 2019), 1–28. https://doi.org/10.1145/3359126

[50] Anbang Xu, Shih-Wen Wen Huang, Brian P. Bailey, and Alejandra Duque-Estrada. 2014. Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-Experts. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing* 24, 74 (2014), 1433–1444. https://doi.org/10.1145/2531602.2531604

[51] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Björn Bjorn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, Vol. 27. Association for Computing Machinery (ACM), New York, NY, USA, 1005–1017. https://doi.org/10.1145/2818048.2819953