

CrowdUI: Supporting Web Design with the Crowd

JONAS OPPENLAENDER, University of Oulu, Finland

THANASSIS TIROPANIS, University of Southampton, United Kingdom

SIMO HOSIO, University of Oulu, Finland

Web design is a complex and challenging task. It involves making many design decisions that materialise preconceived notions of user needs that may or may not be true. In this paper, we investigate supporting the co-design of a website with *visual feedback* elicited from the website’s community of users. Website users can express their needs by re-arranging and modifying the website’s layout and design. To explore and validate this idea, we present *CrowdUI*, a web-based tool that enables members of the community of a website to visually express their design improvement ideas, frustrations and needs, and to send this feedback to the person in charge of designing or maintaining the website.

CrowdUI is validated in a study with 45 users of a popular social media and networking website. Second, our qualitative evaluation with 60 experienced web developers shows that CrowdUI is able to elicit diverse and meaningful feedback. Put together, our results suggest that CrowdUI’s approach constitutes a productive setting for eliciting visual feedback from the user community as a complement to traditional ways of eliciting feedback and participatory design. Finally, based on our experiences, we discuss a design space for crowdsourced web design and provide design recommendations for similar future tools.

CCS Concepts: • **Human-centered computing** → Human computer interaction (HCI); User studies; *User interface toolkits*; • **Information systems** → *Crowdsourcing*.

Additional Key Words and Phrases: crowdsourcing, web design, participatory design, creativity support

ACM Reference Format:

Jonas Oppenlaender, Thanassis Tiropanis, and Simo Hosio. 2020. CrowdUI: Supporting Web Design with the Crowd. *Proc. ACM Hum.-Comput. Interact.* 4, EICS, Article 76 (June 2020), 28 pages. <https://doi.org/10.1145/3394978>

1 INTRODUCTION

Web design is not a trivial task [49]. The design is typically led and informed by only few people (e.g., web designers and corporate stakeholders) who make decisions for many (i.e., the user community of the website). Understanding user needs is imperative for designing an efficient and attractive user interface (UI) [18, 41]. Traditionally, user needs are elicited with questionnaires or face-to-face methods such as interviews or focus groups. While functional, these approaches have limitations. For instance, questionnaires are often multiple choice and do not allow participants to express their needs precisely [34]. Further, these methods are often conducted with small groups of people and therefore lack generality, are costly, and do not scale well [34]. Indeed, “fully understanding customers’ needs is often a costly and inexact process” [52], and novel online co-design methods are increasingly being explored in academia and practice alike.

Authors’ addresses: Jonas Oppenlaender, jonas.oppenlaender@oulu.fi, University of Oulu, P.O. Box 8000, 90014, Oulu, Finland; Thanassis Tiropanis, t.tiropanis@southampton.ac.uk, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom; Simo Hosio, simo.hosio@oulu.fi, University of Oulu, P.O. Box 8000, 90014, Oulu, Finland.

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

Conducting studies remotely online helps avoid common biases in traditional methods. To this end, services such as UserTesting.com have emerged to match some of these needs. But these services come with a monetary cost that is out of reach for many website owners. Further, these services employ extrinsically-motivated testers who may not be familiar with the website to be evaluated. In this paper, we explore how to harness the community of a website, its authentic users during their visit of the website, in the web design process.

Our solution, *CrowdUI*, is a web-based tool that elicits visual feedback on the design through crowdsourcing from the user community of a website. CrowdUI allows for surveying and collecting rich types of design suggestions from the website users, such as moving, deleting, and resizing elements of the website. CrowdUI tackles the problem of gathering feedback for the re-design of a live website at scale by engaging website users to collectively provide suggestions for improvements of the user interface. From the feedback requester’s perspective (typically the web designer or developer), CrowdUI aggregates the design suggestions as convenient heatmaps that depict suggested modifications and thus inform about the design needs and preferences of the community.

We discuss the design of CrowdUI against the backdrop of a design space for crowdsourced web design that exemplifies how web design can be supported with crowdsourcing. Using CrowdUI, we explore four research questions:

RQ1: How can the crowd remotely support the design of a website?

RQ2: How do website users use CrowdUI, and what are their motivations for modifying the UI?

RQ3: How much and what value do website users with different experience (beginners, intermediate or expert users of a website) derive from the tool?

RQ4: What benefits does CrowdUI provide to feedback requesters?

The results of our evaluation of CrowdUI with website users demonstrates the potential of harnessing the community of a website for contributing to the design process with visual feedback. We found CrowdUI to be a highly engaging tool that allows users to visually express their needs. The primary motivation of website users to use CrowdUI was to increase their perceived ease of use of the website and to remove unused UI elements. Both beginner- and intermediate-level website users found value in the tool and the peer-evaluated designs of others. Self-declared experts, on the other hand, did neither want to use their own modifications nor the modifications created by other users. Users with a beginner-level expertise were most open to use the tool again.

We also evaluated the tool from the perspective of the feedback requester. Skilled web developers were able to draw actionable conclusions from the user-created modifications. While some of the skilled web developers struggled to make sense of the individual crowdsourced designs, a clear benefit emerged from the heatmaps which provide an aggregated view into the needs of certain groups of website users. Our approach to aggregating design suggestions demonstrates that visual feedback from non-expert website users may hold value to the designers and developers of a website. Based on our studies with CrowdUI, we formulate design recommendations for crowdsourced website design systems and for eliciting visual feedback in crowd feedback systems, and discuss avenues ripe for future research in this area.

This article is organised as follows. After reviewing related work in Section 2, we present a motivating scenario in Section 3, and explain the design rationale for our crowdsourced website design tool, as well as the design and technical implementation of the tool. The design rationale is presented in form of a design space for crowdsourced website design which answers our first research question (RQ1). In sections 4–5, we present a user-focused study with 45 website users and answer RQ2 and RQ3. To investigate RQ4, we evaluated CrowdUI in two studies with web developers, one focused on interpreting the individual designs created by website users, and one focused on aggregating design modifications in

heatmaps (Section 6). We provide design recommendations and discuss the limitations of our study and future work in Section 7. We conclude in Section 8.

2 RELATED WORK

Supporting creativity in design has been considered one of the grand challenges in Human-Computer Interaction (HCI) [50, 51]. Given the inherent emphasis of crowdsourcing on collecting insights rapidly and inexpensively, crowdsourcing has been suggested as a key approach for supporting creativity [3, 58] and for providing authentic, timely and affordable feedback [13]. While there have been many participatory systems described in related HCI literature (e.g., [2, 4, 8, 9, 26]), we firmly root our work in web design and online feedback systems. As such, our work touches on the topics of crowdsourcing design feedback online, crowdsourcing website design, and aggregating crowdsourced feedback.

2.1 Design Feedback from the Non-Expert Crowd

Yu and Nickerson found that large-scale, crowd-based parallel design processes may successfully be conducted online [59]. Further, prior work found that non-expert members of the crowd are able to provide useful aesthetic preference judgements [19, 44], and a sub-group of workers on online crowdsourcing platforms enjoy and actively seek out creative tasks [43]. In light of this, researchers and practitioners have explored ways of expanding the feedback elicitation from traditional co-located settings to a crowd of people online. *Crowd feedback systems* [32, 56] are computer-mediated systems that enable an individual to collect feedback and critique from a large number of people, including potential end users. These systems provide an alternative for asynchronously sourcing feedback, decision support, and critique from a crowd with a diverse background and knowledge.

Prior work investigated how crowd feedback systems may support designers in their work. Koyama, for instance, sourced aesthetic preference judgements from the crowd to facilitate interactive design exploration [24]. The designs in Koyama’s system are derived computationally and the crowd contributes by tweaking design parameters. *CrowdCrit* [32] provides designers with written critiques from non-expert members of the crowd. *Voyant* [56] is a similar feedback system that elicits structured perception-oriented feedback for graphic designs from a paid non-expert crowd. *Critiki* [16] gathers design critique for crowdfunding campaigns from paid crowd workers. The system’s scaffolded process supports the workers in writing critiques of high quality. Kang et al. enriched individual feedback items with visual examples in their *Paragon* crowd feedback system [21]. The authors found evidence that the crowd provides more specific and actionable feedback with this example-centric approach.

Our system also elicits feedback from the crowd. We focus on eliciting visual feedback from unpaid community members of an operational website. As such, our approach uses *community-based crowdsourcing* [6] or *community sourcing* [20]. The feedback requester in our system is the web designer or the web developer. The crowd consists of members of the website community who provide the feedback requester with visual design suggestions.

2.2 Collaborative and Crowdsourced Website Design

User feedback and participatory design may provide a relevant contribution to the development and design of web pages. In this section, we review a number of systems that use collaboration or crowdsourcing for supporting web design.

Gambit [47] is a collaborative sketching system that involves different stakeholders for sketching user interfaces. *Apparition* [27, 29] is a UI sketching tool in which designers verbally describe changes to the UI to be implemented by crowd workers in real-time. Based on Apparition, Lee et al. presented *SketchExpress* [30], a crowd-powered tool

that engages crowd workers for prototyping interactive interfaces and reusable behaviours. CrowdUI differs in two ways from these systems. First, it uses a high-fidelity website as its canvas, not a low-fidelity prototype or design sketch. CrowdUI thus works at run-time, not design-time. Second, instead of asking the anonymous crowd (in the case of Apparition) or a limited number of co-located stakeholders (in the case of Gambit), CrowdUI aims to engage the community of website users who are likely to be familiar with the website’s design and shortcomings.

Quill [15] is a collaborative design assistant that generates design proposals for the web designer. Quill addresses the problem of providing a consistent user interface across different target platforms with a system in which experts translate requirements into definitions for user interfaces. Our tool, on the other hand, is concerned with website design and evaluation by end users of the website.

The enterprise portal by Akiki et al. [2] is similar to our tool. The system asks enterprise end-users and experts to modify the UI and peer-evaluate user interfaces that were customised by other users. Our tool differs in that it is fully based on Web technologies (HTML, JavaScript and CSS) and not tied to the specific context of an enterprise, although this is a valid use case (see Section 3.1). Further, and more importantly, instead of only configuring a set of features, the website users in our tool can directly modify the layout and visual design of the web page.

CrowdDesign [38] is a platform for crowdsourced web engineering and design. In this system, the anonymous crowd from a crowdsourcing platform is used to create and evaluate web components of a component-based user interface. *CrowdStudy* [40] is another a toolkit by Nebeling et al. for crowdsourcing the usability evaluation of web-based interfaces. As in our system, the crowd is given instructions and guided through a structured process. The crowd, in the case of CrowdStudy, is however merely used to evaluate designs, not create them. A further limitation of this system is its focus on usability measurement and testing, whereas our system could potentially be configured to gather any kind of structured or unstructured feedback from the community of website users.

Of past systems in related literature, *CrowdAdapt* [39] is most similar to our system. CrowdAdapt sources website adaptations from the crowd, but in the context of responsive web design. Our solution differs in that its main purpose is not supporting responsive design or personalisation of the UI for continued use, but sourcing design suggestions from different user segments as a form of visual feedback to inform the design conducted by a web designer.

2.3 Aggregating Design Feedback

In crowd feedback systems, feedback requesters face the challenge of having to explore and analyse a potentially high number of responses. Crowd feedback systems thus call for processing and aggregating of individual feedback items to support the feedback receiver in the related sensemaking activities [14]. Crowd feedback systems address this issue in a number of different ways.

CrowdCrit aggregates feedback from the crowd in a visualisation [32]. Designers can explore top critiques and the feedback given under specific rubrics in the system’s aggregation interface. Feedback in CrowdCrit is, however, only provided in textual form. The *Voyant* system also aggregates feedback for the feedback requesters [56]. In this case, the aggregated feedback includes word clouds and histograms, but also more complex aggregations that include visual elements, such as visual overlays of how elements are perceived in the design. *Zipt* [10] is a tool for conducting crowdsourced remote usability tests of mobile applications. The system collects interaction and design data, and computes aggregated visualisations “that allow designers to quickly understand aggregate interaction and performance metrics” [ibid].

Our tool supports the web designer in evaluating the user-generated design alternatives by aggregating the feedback from the crowd in the form of heatmaps overlaid onto the user interface of the website. A heatmap is a clustering of

observations to highlight interesting patterns [31]. In the context of web-based systems, heatmaps are often used to identify important parts of the website user interface. One benefit of aggregation of data in heatmaps is that heatmaps limit the influence of contradictory or nonsensical edits. The tool, by design, facilitates any kind of edits, and aims to elicit visual feedback from different user groups. These groups can be expected to have different needs. The tool employs aggregation to depict the most desired edits for a given user group, thus essentially leveraging the wisdom of the crowd. The basic assumption is that the majority of the website users will be motivated to make meaningful contributions, thereby limiting the effect of contradictory or inconsistent modifications on the visualisation of results.

In summary, we position CrowdUI as a tool for eliciting visual design feedback on an existing website from the website’s user community. As such, the tool operates in the late-stage design process or run-time of a website. We describe the design and implementation of CrowdUI in the next section.

3 CROWDUI

CrowdUI is a web-based tool for participatory design that harnesses the community of a website through crowdsourcing. CrowdUI invites the users of a website to visually express their thoughts about the website using the website itself as their canvas. Users can directly manipulate (move, resize, and delete) elements of the user interface and peer-evaluate the modified user interfaces created by other users. The purpose of the tool is to provide the website’s designer or developer – the *feedback requester* – with visual design suggestions sourced from the website’s user community. A specific notable aspect of CrowdUI’s feedback elicitation process is the ability to segment website users according to their backgrounds, as different users (e.g., novice and expert users) may have different design needs. While the tool was primarily created to support the feedback requester in eliciting and evaluating user-generated user interfaces, the tool also aims to provide intrinsic value for the website users. Namely, users get a concrete opportunity to contribute to the design itself with their feedback and therefore contribute to common good.

In creating CrowdUI, we faced a number of value-laden decisions and trade-offs. In the following section, we first describe a motivating scenario and then review the decisions made in the design of our tool against the backdrop of a design space for crowdsourced web design.

3.1 Motivating Scenario

Let us consider a scenario that demonstrates CrowdUI’s key features. A web designer at a company is tasked with redesigning a data entry form of the company’s web-based Enterprise Resource Planning (ERP) software. The web designer is located at the company’s headquarters. The end users who access the data entry form are located in the Production Planning departments which are distributed over several different locations. Instead of conducting interviews, focus groups, and workshops with local expert users (associated with considerable costs, if done face-to-face), the designer decides to use the CrowdUI tool to remotely elicit feedback from the distributed users who use the system on a daily basis. The designer suspects that novice and expert users may have different needs. The designer therefore decides to elicit feedback from these two user groups.

Some elements are out of scope of the redesign (such as the company’s logo and the navigation menu). The web designer hence configures these elements to be excluded from any manipulations made by the users. Next, the designer invites the users via e-mail with a broad call to action: “Please modify the data entry page so that it makes you more productive in your daily work.” A number of users are happy to receive this invitation, as there are elements in the UI that have been bothering them for a long time. The users follow the link in the email, read the instructions and familiarise themselves with the tool in an interactive tutorial. The users then proceed to make adjustments to the user

interface and modify the UI to address the shortcomings that they identify. In the next step, users rate a number of interface designs created by other users and leave written feedback.

The designer knows of the long-standing debate about determining the optimal sample size in usability engineering [48]. The web designer decides to follow the original advice of Nielsen [42, 48] and waits until a total of 30 users (15 users in each of the two groups) have completed the task. The designer then takes a look at CrowdUI’s feedback aggregation interface in which the user-provided suggestions are aggregated in heatmaps. The heatmaps as well as the individual design suggestions created by users allow the designer to clearly identify that 1) the submit button is located in a sub-optimal position, and 2) the page contains elements that are not needed for the given task by the users of the ERP system. The written feedback confirms these findings. The web designer can now proceed to adjust the web page’s UI to the needs of the users.

3.2 Design Rationale

The design of our crowdsourced website design tool follows three fundamental principles:

- (1) *Supporting different use cases.* The feedback requester should be able to adapt the tool to a broad range of different tasks and use cases.
- (2) *Usability by non-experts.* User-facing tool kits “must be user-friendly” [52]. To avoid overburdening novices and to make the system easy to use by non-expert users, we decided on a minimal, easily learnable UI [41].
- (3) *Feedback value.* The system needs to present value to the feedback requester. This value proposition must hold true even for feedback generated by users with no design experience and for users who manipulate only very few elements of the user interface.

Next, we discuss the specific trade-offs made during the development of CrowdUI. We derive a *design space for crowdsourced website design* (see Table 1), based on the related literature presented before, and use it as basis for discussing our own design considerations. We wish to acknowledge that, despite our best efforts, this design space may not be exhaustive. On the other hand, we believe that the selected dimensions cover in sufficient detail the most critical decisions necessary to design a system for crowdsourced website design. The design space spans a continuum that answers our first research question (RQ1): how can the crowd remotely support the design of a website?

According to incentive theory, the primary purpose of a participatory web design tool must be aligned with the stakeholders’ motivations. The main stakeholders, in the case of participatory web design feedback, are the feedback requester (e.g., the web designer) and the users of the website. From the users’ point of view, one approach is to support the personal adaptation of user interfaces for continued use (i.e., personalisation, as found in CrowdAdapt). CrowdUI, however, is primarily designed for feedback requesters. CrowdUI is a *crowdsourced website design system that enables the feedback requester to elicit peer-evaluated visual feedback on the design of a website from the website’s user community*. As such, the tool is designed as an “extended workbench” of the web designer. The system enables website users to view the work space of the designer and to make visual design suggestions.

The design rationale entails a number of limitations. We assume that a website already exists. CrowdUI cannot be used to co-design a website from scratch with the help of the user-community, as elements of the website can only be modified, not added. One downside of this decision to position CrowdUI as a tool for summative feedback is that the existing design of the website could potentially influence both the users’ willingness to modify the design and the users’ actual modification of the design. The bias introduced by the quality of design is, however, something that affects all crowdsourcing-based feedback systems.

Table 1. A design space for crowdsourced website design with selected examples for each dimension.

Design dimension	Definition	Examples from related literature
Object of study	Denotes what is being designed (e.g., a website, prototype, wire-frame sketch)	user interface sketch [47] web page [39]
Primary purpose	The main purpose of the system	co-design [47] personalisation [39] feedback elicitation [56]
Stage in the design process	Whether the crowd contributes at an early or late stage in the design process (cf. formative and summative feedback)	early-stage/prototype [47] final design [39]
Motivation of crowd	Denotes how the crowd is incentivised (i.e., whether motivation is extrinsic or intrinsic)	paid [27] volunteer [39]
Type of crowd	Whether tasks are given to an anonymous crowd of online workers, a community of users, or other groups	crowd workers [27] community members [2]
Type of task	The type and complexity of the task given to the crowd	microtasks [27] macrotasks [39]
Collaboration	Whether members of the crowd work on their own or collaborate with other members of the crowd	individual [40] collaboration [15]
Type of collected data	The form of the data collected (e.g., written feedback, user interface manipulations)	visual [39] written [21]
Constraints and operations	The available operations and level of freedom given to the crowd	tweak parameters [24] manipulation of UI [39]
Longevity of data collection	Whether data is collected in an ongoing or one-off crowdsourcing campaign, or through other means	one-off campaign [10] co-located session [47]
Responsiveness of the design	Whether the design adapts to different screen sizes and form factors of devices	responsive [39] static design [56]
Evaluation method	Denotes how the crowdsourced data is evaluated	manual inspection [21] peer evaluation [2]
Aggregation of collected data	How the system supports the designer in making sense of the crowdsourced data	histograms, word clouds, visual overlays [56]

A further design decision concerns how the design is presented to the user. Our solution turns the web page into a static page. CrowdUI reproduces the exact layout as displayed on the web designer’s desktop screen, and only one page can be evaluated at a time. As a consequence, even users on mobile devices are presented the desktop design, and this may make CrowdUI less suitable for mobile use (unless the design is optimised for mobile devices in the first place).

Another design decision is to reproduce the web page as a static page without any possibilities of interacting with the web page. CrowdUI therefore cannot be used to support interaction and user experience design. We further acknowledge that responsiveness (i.e., the ability of a website to adapt to different screen sizes) is an important aspect of the user experience of websites. However, eliciting feedback on a responsive design would result in a confusing amount of different layouts returned to the feedback requester. Individual modifications of the user interfaces would be difficult to interpret in context of the different screen sizes and could negatively affect the sensemaking process of the feedback requester. The main aim of CrowdUI is to capture authentic feedback on the static visual design of a single web page from the website’s user community, and not to provide an authentic experience or personal adaptation of the UI for continued use.

Finally, aggregation of crowdsourced feedback is an important mechanism to support the feedback requester in making sense of the crowdsourced feedback. Feedback could be aggregated and visualised in a number of ways. Voyant, for example, aggregates feedback in histograms, word clouds, and visual overlays of behavioural information onto the

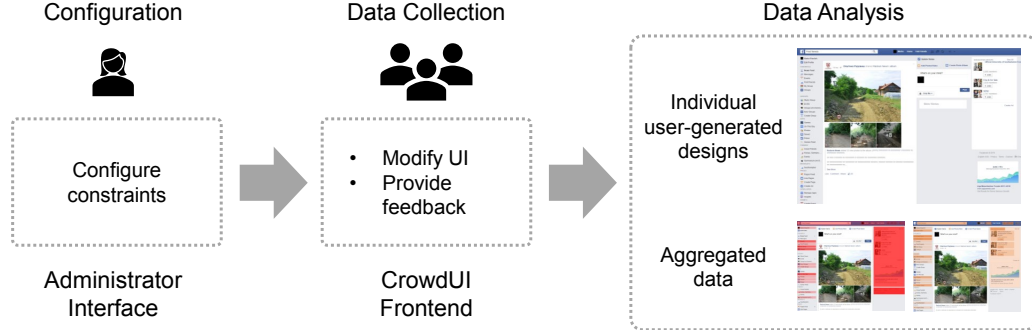


Fig. 1. Overall feedback elicitation process from the perspective of the feedback requester. The figure depicts three phases for requesting feedback. First, the feedback requester configures the elements of the website’s user interface. Feedback is then requested from the website’s users, who can modify the user interface and additionally provide written feedback. The requester then has the option to analyse the user-generated user interfaces individually or in aggregated form as heatmaps.

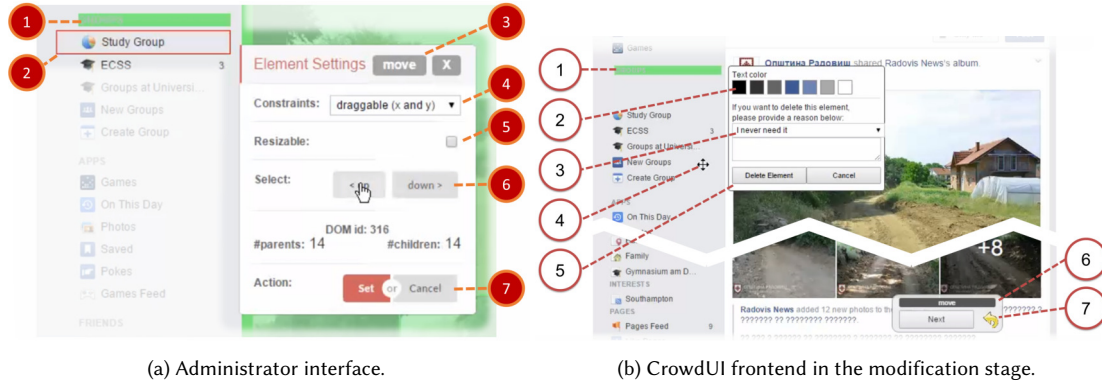








Fig. 2. **Left:** Configuration dialog for grouping and configuring constraints of UI elements in the administrator interface (relevant parts highlighted). The dialog box opens when a DOM element is clicked and is positioned next to the currently selected element ②. The box can, however, also be moved to not obstruct the view ③. The dialog box allows the feedback requester to group DOM elements into logical groups. To group elements, the webmaster can iteratively expand the current selection (highlighted in red ②) by navigating the DOM hierarchy with two buttons ⑥ until the desired grouping is reached. With a drop-down menu ④ and a checkbox ⑤ (see Table 2), the webmaster may apply constraints to the selected element group. The configuration of an element (or element group) is saved with a click on a button ⑦ and the dialog box is closed. Configured elements are highlighted with a green colour ①.

Right: Screenshot of the frontend interface for website users (Stage 4). The user is presented a static view of the web page. When the user hovers the mouse over an element, the mouse cursor indicates the directions in which the element can be moved ④ (in the depicted case, the element can be moved in both directions). Elements can be freely moved (by dragging them to another position) and resized (by dragging handles on the border of the elements). Right-clicking on an element reveals a dialog box ⑤. When the dialog box is open, the currently selected element (or element group) is highlighted in green ①. With the dialog box, users have the option to change the text colour of elements with a predefined colour palette ②. For each deletion of an element, an optional reason is elicited from the user with a drop-down menu and free form text input field ③. The last operation can be undone by clicking an undo icon ⑦. Once the user is satisfied with the changes, the modification of the UI is saved with a click on a button ⑥, and the user is directed to the next stage.

design [57]. We decided on heatmaps as an intuitive and popular means of aggregating and visualising behavioural information on the Web.

Table 2. Constraints that can be configured by the web designer for each element (or element group) and the respective mouse cursor when a website user hovers over the element (or element group).

Constraint	Definition	Input	Cursor
not draggable	Element cannot be dragged	drop-down	
not draggable/deletable	Element cannot be dragged or deleted	drop-down	
draggable (x and y)	Element can be dragged in both directions	drop-down	
only-x	Element can only be dragged horizontally	drop-down	
only-y	Element can only be dragged vertically	drop-down	
resizable	Element can be resized	checkbox	

3.3 System Description

From the requester’s point of view, the process of eliciting visual feedback is structured in three parts: configuration, data collection, and data analysis (see Figure 1).

3.3.1 Configuration. The feedback requester configures the web page in the administrator interface. The feedback requester may group UI elements (see Figure 2a). These groups can then be manipulated by website users as a logical unit. Grouping elements is important for preventing fragmentation of the design and for preventing invisible elements interfering with the manipulation of visible elements on the web page. CrowdUI leverages the Document Object Model (DOM) for deciding on the grouping. Using the DOM hierarchy, children of a DOM element inherit the parent element’s configuration. That means if an element with many children (e.g., a sidebar) is set to be draggable, all children will move along with the parent element and keep their respective spatial distance while being dragged. An exception is the configuration option for resizing elements, which currently is only applied to the parent container, not its children. Using the DOM hierarchy, configuration can also be applied as superimposed layers. All children of a DOM element inherit the parent element’s configuration. Inherited configuration can be overwritten by configuring the child elements individually. A sidebar can, for instance, be configured to be draggable in both directions, whereas the contents of the sidebar can be configured to be only draggable in one direction. CrowdUI supports the designer in determining the desired grouping in an interactive way (see Figure 2a for more details).

The feedback requester can enable five types of constraints for each UI element or group of elements (summarised in Table 2): 1) unconstrained movement of an element (or group of elements), 2) horizontally or vertically constrained movement of an element (or group of elements), 3) resizing elements by dragging the edges of the element, 4) changing the text colour of an element, and 5) deletion of an element (or element group). By default, CrowdUI treats any unconfigured elements as being ungrouped and unconstrained in movement, that is, *draggable (x and y)* is the default option for all unconfigured elements and also the default option in the drop-down menu (see Figure 2a).

3.3.2 Data Collection. Once the web page is configured, the feedback requester invites the crowd – the users of the website – to use CrowdUI, for instance by sending them a link in an e-mail or by adding other calls to action on the web page.

The website users are guided through a six-stage process, summarised in Table 3. The users are first welcomed on a landing page explaining the purpose of the application. The next stage collects demographic data with a questionnaire. In the subsequent interactive tutorial, users are introduced to the different operations of the tool. The tutorial includes a UI element on which users can try out the operations of CrowdUI. If this element is deleted, it automatically appears

Table 3. Six usage stages of CrowdUI for website users.

Stage	Purpose
1: Landing page	Explain purpose of CrowdUI
2: Pre-Task questionnaire	Gather demographic data
3: Interactive tutorial	Familiarise user with CrowdUI
4: Modification	Modify the user interface
5: Peer-evaluation	Rate modifications of other users
6: Post-Task questionnaire	Gather unstructured feedback

again after a short delay to allow for further testing. Once the tutorial is completed, users are directed to the web page where they can move and manipulate elements as configured by the feedback requester (see Figure 2b).

In the following stage, users rate the UI designs created by other users. Users are first given an unobstructed view of the modified UI, and can then on request open a questionnaire that is overlaid on the web page. The overlay can be moved and the overlay is made transparent while being moved, to allow the user to take another look at the web page when filling out the questionnaire. Designs are loaded from the pool of user-generated designs with a confidence-based scoring algorithm. For this prototype, we implemented an algorithm based on Wilson’s score interval [54]. This algorithm transforms the average System Usability Scale (SUS) score of a peer-evaluated design into an “up” or “down” vote (using the score of 68 as the divider). Similar confidence-based algorithms are used on social media sites, such as *Reddit*, for sorting and ranking of user-rated content.

After completing the questionnaire, the user is given the option to continue rating UI designs or to proceed to the final stage. Ratings from other members are not visible to the users, so as to avoid social influences, such as group-think or the bandwagon effect.

The final stage in CrowdUI comprises of an additional survey to solicit unstructured feedback from the users. The data collected in the pre-task or post-task questionnaires may form the basis of a user segmentation to be used in the data analysis, but the two questionnaires can be adapted to serve other purposes (such as, for instance, measuring the pre-task and post-task affective state of the user). In our study, we used the final survey stage to gather open-ended feedback about the usability and perceived value of CrowdUI itself.

3.3.3 Data Analysis. Once feedback is collected, the feedback requester has the option to “replay” modifications and analyse the UI designs individually. Further, CrowdUI visually aggregates the modifications in heatmaps to support the feedback requester in analysing the design suggestions and discovering trends in the collected data. The colour coding in the heatmaps represents the absolute number of users who manipulated the respective UI elements. The darker the colour shading, the more users manipulated the respective UI element. Currently, the system supports two kinds of heatmaps: one for deletions of elements and one for movement of elements. Heatmaps can further be generated for different user groups, based on the user segmentation derived from the pre-task or post-task questionnaire. In our study, for example, heatmaps were generated based on the users’ self-declared expertise with the website resulting in six heatmaps in total (2 types of modification \times 3 user groups).

The main aim of the heatmap visualisation is to give the feedback requester an overall impression of the modifications. The heatmaps are calculated based on the individual modifications collected. The absolute frequency of modifications (deletions and movements) is collected for each DOM element and each user group, respectively. The frequencies are then normalised for each heatmap to a scale ranging from 0 to 1. The resulting values are mapped to shading levels.

We decided to visualise the values in increments of 0.2, that is, there were at most five different shading levels in each heatmap: (0, .2], (.2, .4], (.4, .6], (.6, .8], (.8, 1]. Unshaded parts of the UI represent parts that were not modified by any users, while darker shaded parts were modified more frequently by different users.

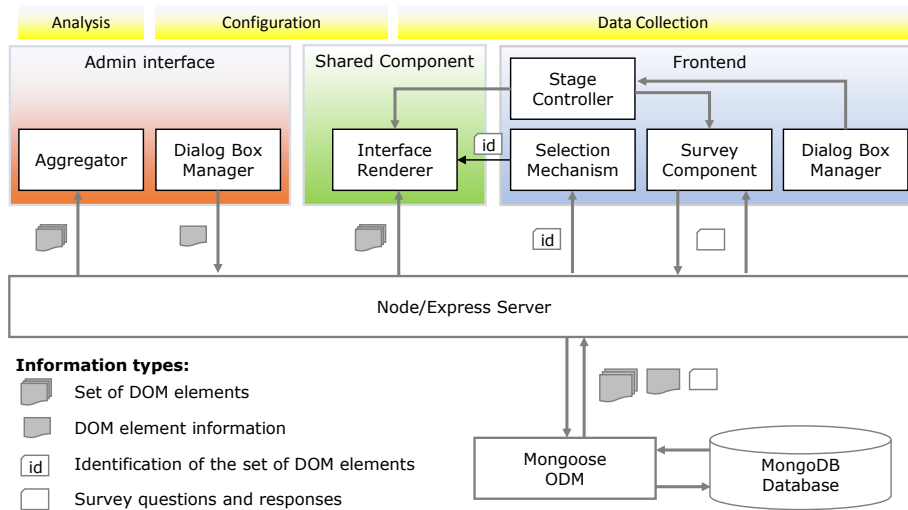


Fig. 3. Components of CrowdUI's architecture and their relation to the feedback elicitation process.

3.4 Implementation

CrowdUI is composed of two sub-systems: the administrator interface for the feedback requester and the CrowdUI frontend for the website users (cf. figures 2a and 2b). The two sub-systems leverage JavaScript libraries which are injected into the host page. Backbone.js is used for synchronising UI modifications with the server, and jQuery UI for manipulating DOM elements. The server is running a REST-less API enabled by Node.js and Express. Data is saved on the web server after each interaction with CrowdUI. Information about each DOM element and the respective modifications made by the website user are stored in a MongoDB database using Mongoose for Object Data Modelling (ODM).

The architecture of CrowdUI is depicted in Figure 3. Both the administrator interface and the frontend make use of a shared component that queries the server for a set of saved element configuration. The component iterates through this configuration and applies CSS and positional information to each DOM element of the web page. A loading bar is displayed while the configuration is being applied. On the frontend, the *Dialog Box Manager* renders the six usage stages and instructs the *Stage Controller* to advance between the six stages. The content of each stage as well as the questionnaire items are defined in JSON objects. The *Survey Component* queries and processes the questionnaires and oversees that all required form fields are filled. In the administrator interface, the *Dialog Box Manager* renders the dialog and sends element configuration for each configured UI element to the server. Heatmaps are generated by the *Aggregator* which overlays coloured semi-transparent boxes onto the respective UI elements.

The DOM element information is derived as follows. During the configuration of elements in the administrator interface, the relative position of each element on the web page is converted into an absolute offset. This enables CrowdUI to exactly reproduce the layout of the user interface as it is displayed on the feedback requester's screen.

Additional cascading stylesheet (CSS) definitions, such as the z-index of elements, are additionally injected into the DOM elements to further increase the reliability and reproducibility of the UI modifications in different browsers and on different screen sizes. Furthermore, all link anchors and onClick handlers in the web page are deactivated.

We pilot-tested the system with the portal page of our institution’s intranet. The system was able to accurately reproduce the layout of the web page, and UI elements could be manipulated as configured.

4 USER-FOCUSED EVALUATION

In a first study, we evaluated CrowdUI from the perspective of website users. The study was administered remotely without supervision as this approach best reflects the intended usage of the tool, mitigates the observer effect, and leads to ecologically valid results.

4.1 Study Setup

We selected the News Feed of *Facebook* (www.facebook.com) as the web page to be manipulated by participants in our study. Facebook is the world’s largest social network, reporting 1.59 billion daily active users on average for June 2019 [12]. Facebook’s design has been stable for years and Facebook’s users can only make a limited amount of changes to the user interface. Given the large user base, we expected most participants in our study to be familiar with Facebook.

The News Feed was prepared for this study as follows. We first downloaded the web page (including the linked images) and removed all embedded and linked JavaScript code from the source code of the document. The approach of using a replica – instead of, for example, a solution based on injection of JavaScript – provisions each participant a similar experience of the website being evaluated. Naturally, the trade-off here is between showing participants the website populated with their own up-to-date content and our replica. We found it more important to provide every participant the same experience from the UI perspective rather than to show participants their own feed. The latter might be distracting or otherwise jeopardise the study in unforeseeable ways (i.e., distractions by the content itself rather than focusing on the user interface). For this reason, we replaced the content of the News Feed with random content that was expected to be geographically distant to the participants (we further explain the reason for this choice in Section 7.1.2).

4.2 Procedure

We recruited participants (i.e., Facebook users) through two channels over a period of 12 days. First, we posted a call to participation in a dedicated Facebook group. Second, we ran an advertisement campaign on Facebook, specifically targeting English speaking Facebook users from the United States, Canada and Australia. Participation in the study was voluntary and not rewarded.

Participants were guided through the study with CrowdUI’s inherent six-stage process as detailed in Table 3. The landing page was modified to include a consent form. Participants were allowed to spend as much time as they wished on the tutorial and the modification of the user interface. Participants were allowed to modify the user interface as they saw fit. No constraints were added to the UI elements in this study, but the UI elements were grouped in logical groups. Participants were asked to create at least one modified design. After creating this design, the participants were allowed to continue creating more designs. The task instructions given to the participants were: “Please adapt the Facebook user interface to your liking.”

4.3 Collected Data and Data Analysis

For quantitatively evaluating the engagement of users with the system, we logged the number of times that the training element was moved in the interactive tutorial, the session length, and the number of elements moved in each created design. For peer-rating the designs (in Stage 5), we employed the System Usability Scale (SUS). While SUS is admittedly not optimal for evaluating static designs, it is suitable for validating CrowdUI's functionality and, more critically to our study, encourages participants to exert thought when rating the modified designs created by other participants.

The qualitative part of our analysis was data-driven and followed a “small q” approach [22]. We analysed and coded all open-ended questionnaire items. We understand ‘codes’ as “comments linked to extracts of text, indicating material identified by the analyst as relevant to their research question” [23]. We inductively and iteratively developed codes from the bottom-up as follows. First, we familiarised ourselves with the data and extracted verbatim statements from the responses. Next, we removed duplicates, merged, and clustered the statements into meaningful groups to identify themes and relationships. Following our “codebook approach” to thematic analysis [7] and reflecting our positivist viewpoint, we report the absolute frequency of responses for each theme identified.

5 RESULTS

In this section, we report the results of our user study. We analyse the level of engagement, the modifications created by the participants, and we answer research questions RQ2–RQ3.

5.1 Participants

In our 12-day deployment, we registered 504 visits to the landing page of the study. Over 30% (186) of the visitors started the first stage of the study. Of these visitors, 138 did not completely finish all stages of the survey and were excluded from further analysis. Three participants completed the study on mobile devices (one iPhone, two iPads). Since CrowdUI is specifically designed for desktop environments, we excluded these three participants from further analysis.

Among the final set of 45 participants, Chrome was the most used browser ($N = 25$), followed by Firefox ($N = 10$), Internet Explorer ($N = 9$), and Safari ($N = 1$). We registered three participants using a macOS desktop computer. The majority of participants, however, completed the study on desktop computers running Microsoft Windows ($N = 42$).

The demographics of the participants (P1–P45; seven identifying as female, none as non-binary) are as follows. The participants had a broad range of educational backgrounds. We counted three completed doctorate degrees, seven Master degrees, 14 Bachelor degrees, nine participants with completed vocational training, 11 high school graduates and one participant with no completed schooling. Most of the participants were from the United States ($N = 25$), followed by the United Kingdom ($N = 7$), Canada ($N = 4$) and Australia ($N = 2$). We registered one participant from each of Brazil, Germany, Greece, New Zealand, Portugal, South Africa, and Spain. The age of the participants was collected in 5-year brackets. The youngest three participants fell into the 20–24 age bracket, the oldest one into the 65–69 bracket ($M = 43.8$ years, $SD = 11.8$ years)¹.

The majority of participants reported visiting Facebook very frequently ($N = 14$) or frequently ($N = 15$). Some participants reported spending up to several hours a day ($N = 13$) on Facebook, others 1 hour ($N = 12$) per day. Many participants rated their own knowledge of Facebook as being “average” ($N = 19$; 42.2%). Almost two thirds of the participants ($N = 28$; 62.2%) rated themselves as being an “intermediate” Facebook user, 11 as “beginner” and six had

¹Note that we used the midpoints of the age brackets to calculate these descriptive statistics which may introduce bias.

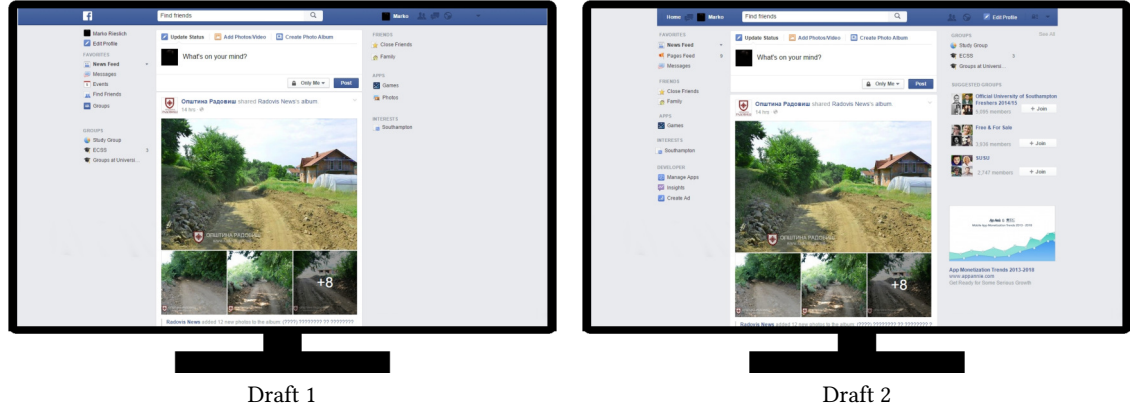


Fig. 4. Two modified user interfaces created by participants in the study. In Draft 1, the participant deleted some elements of the sidebar and navigation menu, including the advertisement and suggested groups. Changes are more subtle in Draft 2, with only certain elements of the sidebar removed and re-ordered.

self-declared “expert” knowledge of Facebook. Approximately half of the participants ($N = 22$; 48.9%) stated they use at least “some” features of Facebook. None of the participants used all features.

5.2 Level of Engagement

To investigate user engagement, we analysed the participants’ behaviour in the modification stage, the peer-evaluation stage, and the tutorial.

Participants created 62 different user interfaces (in the following also referred to as *drafts*) during the study. Participants spent up to 34 minutes on creating their draft ($M = 3.4$ min, $SD = 7.3$ min)². We recorded 200 peer-evaluations in total. Participants peer-evaluated 3.2 drafts on average ($SD = 1.8$, $Mdn = 3$). One participant peer-evaluated 12 drafts.

The training element in the tutorial was moved up to six times ($M = 1.2$, $SD = 1.2$). Half of the participants (51.1%) moved the training element at least once. We did not find a correlation between the number of moves that a participant made in the interactive tutorial and the number of moved elements in the participant’s own drafts ($r = .007$).

We recorded a total of 1,521 individual movements of UI elements. The number of elements that were moved per draft varied greatly ($M = 25$, $SD = 53.8$, $Mdn = 7$). The maximum number of elements moved in a draft was 336, the minimum was 1. The distribution of the number of moved elements in the drafts clearly followed a power law. Few drafts had many moves, and many had few moves. More specifically, the 13 drafts (20.9%) with the highest number of moves constitute 80% of all element moves, and the single draft with the highest number of moves contributed over 20% of all moves. Similar patterns of participation can be observed in other systems online [17].

5.3 User Interface Modifications (RQ2)

The user-generated drafts varied greatly in their appearance. Two examples of drafts are depicted in Figure 4. An inspection of all drafts suggests that self-rated expert-level users were more conservative in their modifications of the UI, making mostly minor changes. Intermediate-level users, on the other hand, embraced the system and were less conservative with their modifications of the UI (see also Figure 6).

²One participant resumed the modification stage after pausing for 18 hours. We excluded this outlier from the descriptive statistics.

Table 4. Summary of modifications to the web page in the user study.

Theme	Frequency	Definition
Minor	36 (58.1%)	Minor modification of the UI, with less than ten elements modified
Major	26 (41.9%)	Modification that constitutes a major change in layout and appearance, with at least ten elements modified
Finished	53 (85.5%)	Modification appears to be complete
Unfinished	9 (14.5%)	Modification appears to be left in an unfinished state
Layout	60 (96.8%)	Re-arrangement of elements of the UI
Design	2 (3.2%)	Modification of the appearance of elements (resizing, colour change)
Aligned	52 (83.9%)	Elements belonging to a logical group (e.g., the sidebar) are aligned
Misaligned	10 (16.1%)	Elements belonging to a logical group (e.g., the sidebar) are not aligned
Demonetization	15 (24.2%)	Banner advertisement was removed
Footer	21 (33.9%)	The footer (including the copyright notice) was removed
Defacing	3 (4.8%)	Logo and/or branding was removed

To develop a better understanding of the modifications made to the web page, we clustered the drafts into classes of similar modifications. This activity resulted in a set of ten basic themes used to describe and categorise the drafts from different perspectives. The themes are summarised in Table 4. The first eight themes are dichotomous pairs, meaning we categorised drafts as either belonging to one or the other category.

Overall, we found that the *layout* of the web page was modified in 60 of the 62 crowdsourced drafts (96.8%), for instance by re-arranging and deleting UI elements. The operations for changing the text colour and resizing elements were used very sparingly by the participants. Only six participants experimented with modifying the *design* of individual UI elements. Of these participants' drafts, we classified only two as concerning primarily the design of the web page, and not the layout.

Twenty-six drafts were *major* modifications, that is, they were considered as a (sometimes drastic) modification of the original user interface, with at least ten UI elements moved or deleted. Participants made *minor* incremental modifications in 36 cases, with only very few elements moved. Nine drafts appear to be left *unfinished*, with modifications being seemingly unsystematic and random, while 53 drafts (85.5%) appeared to be *finished*. Whether these drafts were actually considered as "finished" by the participant could not be determined, but they appeared finished and "ready-to-use" to us. The number of finished drafts is remarkable and hints toward the feasibility of our approach.

We found ten drafts to contain *misaligned* elements. UI elements that were formerly part of a logical group (e.g., the links in the sidebar) were not horizontally or vertically aligned in these ten drafts. One reason for the misalignment is that we did not apply any restrictions to the movement of UI elements in our study. Limiting the movement in one direction or snapping to a grid could be considered to solve this issue. A high number of drafts (52; 83.9%) had all elements *well-aligned* (within a small tolerance), which again is a testimony of the participants being engaged and paying attention to making coherent improvements to the user interface.

In three of the UI drafts, the website was *defaced*, that is, the logo and branding was removed from the web page. The *footer* (including the copyright notice) was no longer visible in 21 cases. Lastly, the *advertisement* was removed in 15 of the drafts. This is a surprisingly low number, considering the negative effect that advertisement may have on the user experience [45].

5.4 Motivations for Modifying the UI (RQ2)

Twenty-six participants left open-ended responses explaining their motivation for using CrowdUI. Participants were particularly interested in increasing the *ease of use* ($N = 11$), *simplifying* ($N = 9$), and *adapting to visual and personal taste* ($N = 6$). P13 (intermediate), for instance, mentioned he wanted to make the UI “*less cluttered and easier to navigate*.” Two participants felt a need to “*make the user interface [their] own*” (e.g., P7, intermediate). Related to usability, one participant mentioned better *learnability*:

If we had an option to adapt FB interface and choose what we want to see, it would be easier to use it on a daily basis (I often feel overwhelmed with information), and it would be easier to teach our parents or grandparents to use FB if we could simplify the interface for them. (P1, beginner)

Six participants *removed unused/unliked features* and five participants wanted to *group functionality*. P20 (intermediate), for instance, wanted to “*prioritize the items that would be useful for me, and delete the ones I don’t use*.” P3 (expert) was motivated by using “*more of the screen width, and grouping functionality together*.” This would allow “*rearranging [elements] to a logical sequence*” (P21, beginner), and making the UI “*more logical*” (P32, beginner). P45 (intermediate) wanted to make the UI “*easier to find the things I want to see*.” Better utilisation of the screen ($N = 2$), removing the ads ($N = 2$), and matching the style of the browser ($N = 1$) were further mentioned as motivations.

One participant (P37, intermediate, age 50–54) mentioned the need for modifying the UI for *accessibility*:

I use disability adaptation technology when I am using my machine. My main reason for changes was to make the screen easier for me to navigate on a physical and visual level. I also wanted to get rid of extraneous screen crap that I don’t and won’t use, such as links to games and apps that do not work with my equipment.

In addition to motivations, we explored what participants with differing levels of expertise in using the website thought about the tool and its different operations in the next section.

5.5 Value of the Tool for Users with Differing Expertise (RQ3)

In the post-task questionnaire, we inquired about the perceived value of the tool for the participants with four questions on a five-point Likert scale (see Figure 5). We asked the participants about the perceived usefulness of the tool (“This tool would be useful for me”), their willingness to reuse the tool (“I would like to use this tool again”), whether participants would want to use their own modified UI (“I want to permanently use the adapted interface that I created”), and whether participants would want to use the designs created by other users (“I want to permanently use one (or several) of the interfaces that were presented to me”). As explained earlier, the participants’ level of expertise (that is, familiarity with Facebook) was elicited by asking the participants to self-declare their knowledge of the website (on three levels: beginner, intermediate, and expert) in the pre-task questionnaire.

The results depicted in Figure 5 suggest that beginner- and intermediate-level participants responded more positively than expert-level participants. Among the three user groups, beginner-level participants were most open to experiment with the tool. Sixty-four percent of the beginner-level participants were willing to use the tool again, opposed to only 18% who did not want to do so. One reason for the latter is that the tool may have overwhelmed beginner-level users. Three beginner-level participants specifically mentioned being “*overwhelmed*” and “*intimidated*” (P9) by the tool and the freedom afforded by the different operations, as exemplified by this quote from participant P8:

With limited interface knowledge, I was a bit intimidated ‘poking’ around. But I did see benefits that could make much more.

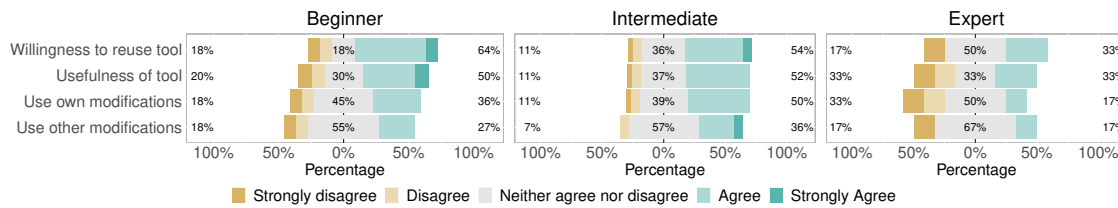


Fig. 5. Evaluation of CrowdUI by beginner-level, intermediate-level, and expert-level users. The percentages on the left, middle and right indicate disagreement, neutrality and agreement, respectively. For example, 64% of the beginner-level users indicated that they agreed or strongly agreed to wanting to use the tool again, while 18% disagreed or strongly disagreed to this statement.

Intermediate-level participants also derived value from the tool. Over half of the intermediate-level participants (54%) agreed or strongly agreed to wanting to use CrowdUI again, compared to 11% who disagreed or strongly disagreed to this question. Of this user group, only 11% thought the tool was not useful. Intermediate-level participants were also more likely to want to use their own designs and the designs created by other participants, compared to the two other groups of participants.

Among the self-rated expert participants, only 17% wanted to use the interface they had created themselves or the interfaces created by other users (see Figure 5). P4, for instance, complained that “*a couple of times the elements such as photos or the menu bar across the top didn’t align properly.*” Two of the participants in this user group demonstrated a strong status quo bias. P5, for instance, expressed that “*if it’s not broken don’t fix it. In other words: leave it alone.*”

6 EVALUATION BY WEB DEVELOPERS (RQ4)

After the user study, we investigated the usefulness of the individual drafts as well as the usefulness of the aggregation of design suggestions in heatmaps (RQ4) in two studies with web developers recruited from the online platform *Prolific* (www.prolific.co). Prolific is a crowdsourcing platform designed for behavioural, user, and market research. In both studies, we required the participants to have knowledge in four areas highly specific to web design and web development: 1) UI design, 2) responsive design, 3) HTML5, and 4) CSS3. In the beginning of each survey, we instructed (i.e., primed) the participants to imagine they were working as a web developer at Facebook and tasked to adapt the design of Facebook’s News Feed to the needs of different user groups. The first study focused on the *individual* designs created by the users in the user-facing study, while the second study focused on the heatmaps as a core feature of CrowdUI. The heatmaps depicted the *aggregated* suggestions made by each group of website users.

6.1 Study 1: Examining Individual Designs

We recruited 12 participants (W1–W12; 3 female; aged 20 to 46 years; $M = 33.5$ years, $SD = 6.7$ years) with web development experience ($M = 8$ years of experience, $Min = 1$ year, $Max = 24$ years, $SD = 7.1$ years). The study took on average 8 minutes and 6 seconds to complete (amounting to an equivalent hourly pay of UK £10.00).

In a web-based questionnaire, we asked the participants to provide their thoughts about a sample of eight UI drafts created in the user study. We hand-selected the drafts to showcase a representative and diverse set of possibilities that the users came up with (in the earlier, user-focused evaluation). In this study, we did not reveal from which user group (beginner, intermediate, or expert) the draft originated from.

Seven participants (58.3%) agreed that the individual drafts generated by the crowd would be useful in adapting the Facebook News feed to better match user needs. These participants found merit in seeing the visual design suggestions

from the users, as they provide guidelines for designing the UI, “a perspective on what the actual user would find useful” (W7), and “concrete ideas and opinions from users as actual design” (W11). The tool was seen as useful in providing a comparison between “various UI designs to find the best design for its intended purpose” (W10). The approach seemed to be “well suited to crafting a more focused or tailored experience for different user groups” (W3).

Two participants neither agreed nor disagreed to whether the tool would be useful, one disagreed, and two strongly disagreed. The participants who were undecided expressed confusion with the UI modifications (W4, W6). Certain drafts were seen as “janky” (W6) and “confusing in the layout” (W4). Participant W4 broached the issue that some UI drafts looked very similar to Facebook’s original design, with only few modifications made. Two participants raised issues related to the skills and ability of regular users, noting that “users are not designers” (W7):

I like the idea of people giving their own opinions about the UI, but i don’t think that is useful for a web developer as a tool because of the inexperience of such users (beginners and intermediate) in UI creation. (W6)

The visual design was also seen in a critical light by W3:

[The tool] doesn’t address other important elements like visual design. Simply eliminating large portions of the newsfeed won’t necessarily provide beginners with a more user friendly experience – at least not on its own.

Half of the participants said they would like to use the tool again, three neither agreed nor disagreed and the same three participants who disagreed earlier also disagreed or strongly disagreed with this question. Among those who agreed, five participants mentioned that the tool provides useful guidelines for the development and improvement of the UI. W1 was undecided whether he wanted to reuse the tool. While the tool “worked good,” W1 noted that “every draft seems to adapt in a different way for certain kind of people.” Among those who disagreed or strongly disagreed, W5 was not convinced of the usefulness of the tool, since the modifications looked “amature [sic]” to him.

As a particularly insightful anecdote, W12 suggested that users should have to try out their own drafts for a few days and continue modifying the UI “until she/he gets something that he/she really likes. The final version of the UI is the one that should be studied.”

In summary, and in relation to providing insights to RQ4, our first study revealed that while some of the individual crowdsourced drafts may hold value for the feedback requester, others may be confusing and difficult to evaluate and interpret because, for instance, only few elements were modified in the draft or the draft was left in a seemingly unfinished state. This confirmed our initial hypothesis, and in a more substantial study we looked into visually aggregating the crowdsourced feedback as a means to assuage these concerns.

6.2 Study 2: Aggregating Design Suggestions

We recruited 48 participants from Prolific with the same qualification criteria and study setup as earlier. The survey was released in batches of 12 in the morning, noon, afternoon and evening. Participation took on average 11 minutes and 54 seconds (resulting in an equivalent average hourly pay of UK £9.00). The majority of the participants (V1–V48; 11 female; aged 18–56 years, $M = 30.9$ years, $SD = 8.4$ years) were experienced in web development ($M = 7.4$ years of experience, $Min = 1$ year, $Max = 23$ years, $Mdn = 5$ years, $SD = 5.4$ years). The participants were overall well familiar with Facebook (only three of them rated themselves as having little to no familiarity with Facebook). The majority of the participants was in full-time employment (28 participants), followed by part-time employment (9 participants). The sample contained 14 students.

The study was conducted with a web-based questionnaire and the procedure was as follows. After consenting to the study and a short briefing, participants were asked to comment on six aggregated website modifications (heatmaps).

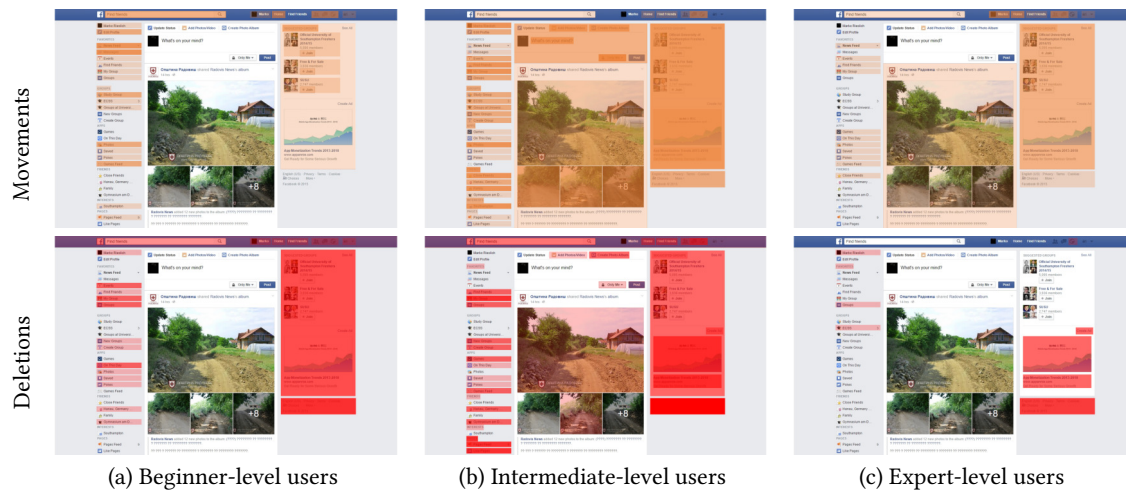


Fig. 6. Aggregated movements (top) and deletions (bottom) of UI elements by self-rated a) beginners, b) intermediate, and c) expert users. The darker the shading in the heatmap, the more users moved or deleted the respective element.

Participants were displayed two heatmaps per user group (beginner-, intermediate- and expert-level users). One of the heatmaps indicated the elements that were moved by the respective user group in the user-facing study (see the screenshots at the top of Figure 6). The other heatmap indicated the elements that were deleted by the respective user group (see the screenshots at the bottom of Figure 6). We inquired about the awareness of user needs generated by the heatmaps, whether the participants would use the insights generated by the heatmaps in their own design, and whether the participants would like to use the tool again. We further asked the participants to elaborate on the conclusions drawn from the heatmaps, why (or why not) they would use the suggestions provided by the users, and what role the different user groups played in their thought process. In this study, the user group associated with the heatmaps was revealed to the participants.

Next, we highlight key insights concerning the specific research question (RQ4) set earlier.

6.2.1 Insights into User Needs. The heatmaps provided an *insight into the user needs* (31 participants, 64.4%). A majority of the participants (40 participants, 83.3%) agreed that their awareness on the user needs concerning the user interface increased after studying the heatmaps. Eight of these participants (20%) strongly agreed to this statement.

Two key categories of insights from the heatmaps emerged. The heatmaps supported the participants in learning more about the *user groups* ($N = 35$) and about the *user interface* ($N = 27$).

Among the participants who had insights into the user needs of different user groups, 29 commented on the heatmaps generated by experts, four on the intermediate-level heatmaps, and 13 on the beginner-level heatmaps. Nine participants commented on the *commonalities between the user groups* and noticed that *different user groups have different needs* ($N = 4$). The heatmaps allowed the participants to form theories about the different user groups. V23, for instance, found the feedback useful in understanding the different user groups:

Expert-level users will most likely use the website much more than the others, but they are also more likely to have adapted to the website as it currently is, slightly blinding them to possible improvements. Beginner-level users show what new people expect and want when coming to the website, but they are less likely to know

about the secondary features. Intermediate users are the best of both worlds, but are also likely to cover a larger group of users who all want slightly different things from the website.

By far the most common observation regarding the user interface was that experts appeared to make fewer changes to the UI ($N = 18$). Eight participants speculated that *experts use more features*, and that experts are better able to ignore content ($N = 1$), more accurate ($N = 1$), or know the UI better ($N = 1$):

The expert knows the little things and what can be done to improve the interface without having to make many changes at once. (V20)

Experts may also be “*more used to the interface, so their changes are not that many compared to the beginner/intermediate users*” (V2). V19 stated that there may be “*issues with the design that beginners don’t notice but that experts manage to accommodate/work around*” and that experts may have “*developed coping mechanisms*” to accommodate these issues. In making sense of the edits done by users, five of the participants speculated that beginners perceived the UI as *cluttered, confusing, and overwhelming* and preferred a more *simplistic UI* ($N = 4$) due to being “*confused with too many options*” (V38). One participant attributed these observations to the UI having a “*learning curve*” (V19).

Regarding the user interface, the tool allowed identifying *important and unimportant parts of the UI* ($N = 23$), and potential candidates of parts of the UI that could be adapted for different user groups. For instance, 12 participants identified the *sidebar* as an unused or disliked part of the UI by beginner-level and intermediate-level users. One participant suspected that only experts make use of the information in this sidebar (V32). V48 suggested to expand the News Feed and convert the layout to a two-column layout for beginners and intermediate users.

6.2.2 Usefulness of the Aggregated Feedback for Web Design. Given their assumed role in the study, a vast majority of the participants found the insights generated by CrowdUI’s heatmaps useful and were willing to incorporate the findings in their own design (12 strongly agreed (25%), 31 agreed (64.6%), four neither agreed nor disagreed (8.3%), and one strongly disagreed (2.1%)).

Involving the user in the decision making was one of the main *motivations* of the participants for including the findings in their own design. Over one third of the participants (20 participants; 41.7%) explicitly mentioned the tool would support their *user-centred approach* to web development. Twenty-eight participants mentioned the tool would *support the UI design*:

In my opinion there is nothing better than a feedback from a real user. Because they frequently engage (in this case news feed) and interact with the news feed, they have the best understanding of what they would like to change. It’s something i would definitely consider when improving my work. (V29)

Sometimes you get so ingrained in what you think is best that you forget what the experience for the everyday user is like. I would think it’s important to use a tool such as this periodically to make sure things are on the right track. (V40)

The tool was seen as providing an actionable view on the user needs, allowing to identify parts of the UI that hold little value to certain user groups ($N = 8$):

Qualitative data such as these provide useful insights from which hypotheses can be formed and tested. (V38)

V29 commented that “*it gives clear visual presentation of what bothers people on the news feed. With combined data from all three groups you can make a more wholesome finished design.*” One participant (V43) saw the heatmaps as a communication device to spur “*discussions amongst other developers.*”

Two strategies emerged from the qualitative data. Eleven participants explicitly mentioned they would consider the heatmaps to design for commonalities between the user groups. V48 noted that he “*would take the common insights from all groups and come up with a design that accommodates those insights,*” which V32 noted would provide “*a good balance between functionality and usability.*” About an equal amount of participants ($N = 10$) explicitly mentioned they would create designs for specific user groups. V28, for example, would use the heatmaps to “*design features based on user group,*” and V36 mentioned he would “*create an expert mode, for the advanced ones.*”

Three participants wanted to have more information to support their decision-making, such as the “*direction of movement*” of elements ($N = 2$) and the background of the users ($N = 1$). One participant mentioned that several rounds of feedback elicitation would be required to validate design decisions:

Any time a change is made to the interface, no matter how big or small, heatmaps could almost instantly show you what works and what doesn't, allowing you to adapt accordingly. (V23)

As a general trend, the heatmap of deletions was perceived as more useful than the heatmap of moves. As V17 put it, the “*heatmap for movement seems less useful because you don't know where they moved [the element] to.*”

The majority of participants ($N = 30$, 62.5%) agreed that they would use the tool again. Fourteen participants strongly agreed (29.2%) to this question, four (8.3%) neither agreed nor disagreed and none of the participants disagreed.

7 DISCUSSION

Designing websites is difficult. A web designer needs vision and talent in crafting websites that are both compelling and usable. Yet, there clearly is merit to not only consulting lead users in the participatory design process, as advocated for instance by von Hippel in [53], but involving the website user in a participatory role that is further-reaching than just giving post-hoc feedback. This is much in line with literature on joint application design [55], user-centered design [1, 11], participatory design [37], and co-creation [46] that all involve the user in some capacity during the design and evaluation of products, often to improve the end result in some way, such as extracting user needs or requirements.

With CrowdUI, we are interested in visual feedback (i.e., modifications of the canvas of an existing web page), and the usefulness of these suggestions to the feedback requester. The results from our studies with CrowdUI suggest potential in harnessing the user community of a website in providing visual feedback to feedback requesters (web designers or developers). The quantitative review of the interactions with the system indicates a high level of engagement of the website users with the tool.

In our user-focused study, we found most modifications concerned only the layout of the web page and the design of individual elements was only changed in few drafts. This finding hints towards CrowdUI being primarily useful as a tool for modifying the layout of websites, as opposed to modifying the visual design. Further, participants primarily used the tool to make improvements to the usability of the web page, adapt the page to their liking, and declutter the web page by removing unused elements. Similar motivations have been found in related studies, for instance on “bloated” user interfaces [35]. Accessibility, on the other hand, while of critical importance to a minority of the participants, overall was little reflected in the responses. Yet, accessibility is an increasingly critical requirement in crowdsourced design tools, as website users may not have the knowledge required to produce user interfaces that adhere to accessibility standards, such as the Web Content Accessibility Guidelines (WCAG). CrowdUI, on the other hand, is positioned as a tool for eliciting visual suggestions from the crowd. The web designer is ultimately responsible for implementing the suggestions, using them as guidelines of what people imagine of a UI.

Both self-declared beginner- and intermediate-level users found value in the tool and the peer-evaluated designs of others. Users with a beginner-level expertise were most open to use the tool again. Self-declared experts were not inclined to use their own modifications nor the modifications created by other users. This may be explained with the experts being more picky about the visual design. The status quo bias demonstrated by this user group may further explain this finding.

Regarding the aggregation of visual feedback in heatmaps, we noticed that while some of the skilled web developers struggled to make sense of the individual drafts, a clear benefit emerged when we evaluated the aggregated visual feedback. The heatmaps allowed the participants to identify both commonalities and differences between the needs of the different user groups. We argue that heatmaps are a way of supporting the feedback requester in making sense of the visual feedback from the crowd. The quantitative information encoded in the heatmaps allows the feedback requester to establish the order of importance of user needs, and to identify areas of the UI that need improvement. This value proposition inherently holds true, even if only few elements in the draft are changed by individual website users. Indeed, the experienced web developers in our study reported that their awareness of the users’ needs increased due to the heatmaps and participants were able to identify many areas for improvement of the website’s layout.

Our qualitative evaluation was conducted with authentic website users and 60 participants with 7+ years of experience in web development, making the participants well-knowledgeable of the domain and many of them capable of articulating expert qualitative feedback. The results obtained from both users and experts allow us to confidently articulate some key observations as higher level design suggestions for crowdsourced website design systems similar to ours.

7.1 Design Recommendations

Based on our findings and experiences in our study, we distil three recommendations for crowdsourced website design systems and for eliciting visual feedback in crowd feedback systems.

7.1.1 Collecting and Improving User-Created Drafts. First, many website users moved only very few UI elements in our study, while others abandoned the modification stage altogether and left the draft in an unfinished state. Therefore, it is imperative to distinguish between real suggestions and unfinished drafts, and simply excluding people who did not take a post-task survey is not sufficient. People might want to skip the survey but leave their ideas on the table. A possible solution is to ask users to indicate if their draft is finished and ready for review by other users. As an idea for future systems, such drafts could be returned into the pool of available drafts for other users to “fork” or iterate on.

Website users may not fully be aware of their needs “until they try out prototypes to explore exactly what does, and doesn’t work” [52]. Hence, another idea for potentially improving the tool, as also pointed out by one of the participants in our study, would be to require the website users to use their own modified UI for a certain amount of time. This would most likely lead to more refined visual design suggestions.

7.1.2 Decoupling the Evaluation of the UI from the Content. The subjective evaluation of a website is not only influenced by aesthetic factors, such as the design and layout, but also by the content of the website. This conflation with content [5] is to be avoided, because the interaction of website users with the content of the website during the peer-evaluation could potentially bias the results. A user could, for instance, become distracted or decide to rate a draft poorly based on its content, not its design or layout. There is a compromise to strike here. In our case, we managed the decoupling by making a static replica of the News Feed and replacing the content with less emotionally charged content. This content was taken from a News Feed of a municipality in a rural area and therefore considered as geographically distant to the feeds that our study participants typically follow. Such a strategy is most likely not feasible on live websites.

However, an equally good strategy will be simply to collect more feedback. The quantity will account for the variances of individual users that will be affected by the content currently shown on the website.

7.1.3 Aggregating Design Suggestions for Decision Support. While the individual suggestions made by the users hold some value for the feedback requester, evaluating each individual change in user-generated design suggestions would be a work- and time-intensive process. The web developers in our second study clearly articulated that the aggregated information per user group (in our case self-declared beginners, intermediate, and expert Facebook users) holds distinct value to them. The visual aggregation of user-provided suggestions supports the synthesis into a final design. User-provided suggestions are inherently subjective, however, and there is no guarantee that they converge into a single version of the user interface for a given user group or a given design task. Therefore, a heatmap like in our system may be most useful in supporting designers to gain awareness of what type of changes a specific group of users would like to see, as a starting point for further investigations.

7.2 Limitations

We acknowledge several limitations of our study. First, the study setup was not entirely authentic from the participant's perspective. Each participant was shown the same web page. Yet, this was mandated by the goals of our study: investigating CrowdUI itself and not improving Facebook *per se*. This design choice allowed us to ensure equal conditions for all participants in the study.

Second, we note that for deployment on a website, a more purposeful task should be given to the users. Yet, for evaluating the system, we argue that a general task (modifying a familiar news feed) was a good match. We acknowledge task diversity as a future step with CrowdUI.

Third, we explored the user groups in a slightly sub-optimal manner. Classifying website users into groups by their self-reported expertise may not lead valid results [33]. However, our approach is suitable for demonstrating the usefulness of CrowdUI and validating its approach. Further, the pre-task and post-task questionnaires can be configured to accommodate any kind of questionnaire that could be used to derive a different type of user classification.

We also acknowledge that responsiveness (i.e., the ability of a website to adapt to different screen sizes), is an important aspect of modern websites. However, sourcing feedback on a live website with a responsive design (as opposed to a static design snapshot) would result in a confusing amount of different layouts returned to the feedback requester. This would make it complicated for the feedback requester to interpret and make sense of the different modifications in context of all elicited feedback. This is a future data visualisation problem, yet to be exhaustively addressed by the community.

Last, in our studies we did not exploit the peer-feedback feature of CrowdUI to its full potential. It could have been used in a more beneficial fashion than administering a SUS questionnaire that, in the end, provides scores that have no real validity in this case. Yet, our purpose was to demonstrate that the system is feasible. The SUS made the participants think about the designs more carefully and it allowed us to validate that users can be engaged in giving peer feedback. For this reason, our choice was adequate.

Despite the limitations, and given the extent of our studies with over 500 visitors, 62 collected designs, and feedback from 60 participants with several years of web development expertise, we argue that our findings are sufficiently reliable to warrant outlining reasonable research directions for future work in the space.

7.3 Future Work

In this paper, we tackled the problem of how to engage a crowd of people in collectively improving a web page by eliciting visual feedback from the website community. While we demonstrated the technical feasibility in a study with 45 website users, the technology is yet to be tested on real websites at scale. Crowdsourced web design presents unique technical and social challenges that will need to be studied and resolved to deploy tools like ours on real-life websites. In this section, we discuss key avenues for future research.

7.3.1 Potential Use Cases. CrowdUI is a tool for eliciting visual feedback from the website’s user community. Based on our experiences, CrowdUI’s approach can be helpful in a number of other scenarios, such as:

- remote usability testing,
- personalisation of the UI to different user groups,
- eliciting feedback on specific elements of the UI,
- inspiration and creativity support, and
- comparative evaluations of user interfaces.

But other ends of the design space equally hold value and could be explored in future work. For instance, the technology could be built into a browser extension to allow harvesting feedback for any website. This would enable the collection of distributed design feedback at scale. Kumar et al.’s work on design mining [25] and Moore’s work on design exploration [36] highlight that remotely collecting and exploring crowdsourced design feedback in this fashion could be a valuable avenue for future research. These uses cases, and others, could be explored in future work.

7.3.2 New Ways of Aggregating Visual Feedback. While the heatmaps provided the participants with actionable insights, we acknowledge that in their current form, CrowdUI’s heatmaps only give a qualitative insight into the users’ suggestions. The heatmaps, for instance, do not indicate where elements were moved to, nor do they provide insights why the users moved elements. Improving the aggregation mechanism is an area ripe for future research. Providing heatmaps with rich annotations (e.g., of the reasons why elements were moved) would allow the feedback requester to interactively explore and understand the suggestions made by the users.

7.3.3 Incentivising Users. What incentives are most effective in motivating website users to participate in the web design process in a committed fashion? The crowdsourced web design approach most likely worked in our study because the participants had an apparent interest in changing the user interface. Most of the study participants did only use a small amount of the web page’s features, so removing the unused parts constituted an inherent improvement in usability for these users. This may, however, not be true for every web page and every group of users. Future work must investigate the types of incentives that work best and across websites with a range of different purposes.

7.3.4 Comparison with Other Forms of Feedback. Future work can also investigate a number of questions related to visual feedback. How does crowdsourced visual feedback compare with traditional forms of feedback? How do we combat the inevitable noise in crowdsourced visual feedback (i.e., data quality issues)? How does visual feedback from the intrinsically-motivated community of website users compare to feedback elicited from an extrinsically motivated crowd workers from paid crowdsourcing platforms?

7.3.5 Supporting the Non-Expert Crowd in the Design. Visual aesthetics of websites are a strong determinant of user satisfaction [28]. The typical website user is most likely not trained in design. In this paper, we primarily focused on

supporting the feedback requester in eliciting and making sense of the feedback received from website users. Future work could investigate how to technically support and guide the website users in the visual design of the web page, for instance by using different types of rubrics or quick tutorials on design heuristics, thereby supporting the users in providing useful feedback. Further, applying design patterns to user modifications is an interesting area for future research, while it is also a limitation in our current system in which website users are free to modify the web page as they see fit.

8 CONCLUSION

We presented CrowdUI, a web-based tool for involving the users of a website in the web design process. The tool engages members of the user community of a website to visually express their needs and recommend revisions to the layout of websites. Members of the website's community use the web page itself as a canvas to express their needs by manipulating elements of the user interface.

We investigated the tool in three studies in the context of the Facebook News Feed. The studies with CrowdUI suggest that the participatory approach is feasible, with dozens of design suggestions harvested in our evaluation of the tool with authentic web users. While some of the experienced web developers in our study found the individual design suggestions difficult to interpret, the aggregated visual feedback (heatmaps) generated by the tool allowed to draw actionable conclusions.

Our findings verify the CrowdUI concept as appealing to feedback requesters and to website users alike. The structured process of extracting and peer-evaluating design suggestions from different user groups enables participatory visual design feedback from website users. Feedback harvested from members of a website community, who are likely to be familiar with the site's shortcomings, is a promising area of research. Such feedback is expressive and hence meaningful, valuable, and actionable and constitutes an excellent complement to traditional ways of collecting participatory design feedback. Gathering user feedback at scale for incorporation into live websites could be of great value for the design and engineering of interactive systems. This work is one step into that direction, allowing a website's user community to use direct manipulation to express their frustrations and needs to the designers.

9 ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their helpful and generous comments on the manuscript. This work was in part funded by the Academy of Finland (Grants 313224-STOP, 320089-SENSATE, 316253-SENSATE and 318927-6Genesis Flagship).

REFERENCES

- [1] Chadia Abras, Diane Maloney-Krichmar, and Jenny Preece. 2004. User-Centered Design. In *Berkshire Encyclopedia of Human-Computer Interaction* (2nd ed.), William S. Bainbridge (Ed.). Berkshire Publishing Company, Ltd., Great Barrington, MA, USA, 763–768.
- [2] Pierre Akiki, Arosha Bandara, and Yijun Yu. 2013. Crowdsourcing User Interface Adaptations for Minimizing the Bloat in Enterprise Applications. In *Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '13)*. ACM, New York, NY, USA, 121–126. <https://doi.org/10.1145/2494603.2480319>
- [3] Salvatore Andolina, Hendrik Schneider, Joel Chan, Khalil Klouche, Giulio Jacucci, and Steven Dow. 2017. Crowdbboard: Augmenting In-Person Idea Generation with Real-Time Crowds. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (C&C '17)*. ACM, New York, NY, USA, 106–118. <https://doi.org/10.1145/3059454.3059477>
- [4] Goetz Botterweck. 2011. Multi Front-End Engineering. In *Model-Driven Development of Advanced User Interfaces*, Heinrich Hussmann, Gerrit Meixner, and Detlef Zuehlke (Eds.). Springer, Berlin/Heidelberg, Germany, 27–42. https://doi.org/10.1007/978-3-642-14562-9_2
- [5] Margaret M. Bradley and Peter J. Lang. 2007. The International Affective Picture System (IAPS) in the Study of Emotion and Attention. In *Series in Affective Science. Handbook of Emotion Elicitation and Assessment*, James A. Coan and John J. B. Allen (Eds.). Oxford University Press, New York, NY,

- USA, 29–46.
- [6] Marco Brambilla, Stefano Ceri, Andrea Mauri, and Riccardo Volonterio. 2014. Community-based Crowdsourcing. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. ACM, New York, NY, USA, 891–896. <https://doi.org/10.1145/2567948.2578835>
 - [7] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*, Pranee Liamputtong (Ed.). Springer, Singapore, 843–860. https://doi.org/10.1007/978-981-10-5251-4_103
 - [8] Andrea Bunt, Cristina Conati, and Joanna McGrenere. 2007. Supporting Interface Customization Using a Mixed-initiative Approach. In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*. ACM, New York, NY, USA, 92–101. <https://doi.org/10.1145/1216295.1216317>
 - [9] Javier Caminero, Mari Carmen Rodríguez-Gancedo, Jean Vanderdonckt, Fabio Paternò, Joerg Rett, Dave Raggett, Jean-Loup Comeliau, and Ignacio Marin. 2012. Multidimensional Context-Aware Adaptation of Service Front-Ends. In *Context-Aware Adaptation of Service Front-Ends 2012. Proceedings of the Workshop on Context-Aware Adaptation of Service Front-Ends (CASFE 2012)*, Francisco Javier Caminero Gil, Fabio Paternò, and Jean Vanderdonckt (Eds.). CEUR-WS, 5 pages.
 - [10] Biplab Deka, Zifeng Huang, Chad Franzen, Jeffrey Nichols, Yang Li, and Ranjitha Kumar. 2017. ZIPT: Zero-Integration Performance Testing of Mobile App Designs. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 727–736. <https://doi.org/10.1145/3126594.3126647>
 - [11] Stephen W. Draper and Don A. Norman. 1986. *User Centered System Design*. Lawrence Erlbaum Associates, Hillsdale, NJ, USA.
 - [12] Facebook, Inc. 2019. Facebook Reports Second Quarter 2019 Results. Press release, 24 July, 2019.
 - [13] Eureka Foong, Steven P. Dow, Brian P. Bailey, and Elizabeth M. Gerber. 2017. Online Feedback Exchange: A Framework for Understanding the Socio-Psychological Factors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4454–4467. <https://doi.org/10.1145/3025453.3025791>
 - [14] Eureka Foong, Darren Gergle, and Elizabeth M. Gerber. 2017. Novice and Expert Sensemaking of Crowdsourced Design Feedback. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 45:1–45:18. <https://doi.org/10.1145/3134680>
 - [15] Vivian Genaro Motti, Dave Raggett, Sascha Van Cauwelaert, and Jean Vanderdonckt. 2013. Simplifying the Development of Cross-Platform Web User Interfaces by Collaborative Model-Based Design. In *Proceedings of the 31st ACM International Conference on Design of Communication (SIGDOC '13)*. ACM, New York, NY, USA, 55–64. <https://doi.org/10.1145/2507065.2507067>
 - [16] Michael D. Greenberg, Matthew W. Easterday, and Elizabeth M. Gerber. 2015. Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (C&C '15)*. ACM, New York, NY, USA, 235–244. <https://doi.org/10.1145/2757226.2757249>
 - [17] Eszter Hargittai and Gina Walejko. 2008. The Participation Divide: Content Creation and Sharing in the Digital Age. *Information, Communication & Society* 11, 2 (2008), 239–256. <https://doi.org/10.1080/13691180801946150>
 - [18] Marc Hassenzahl. 2004. The Thing and I: Understanding the Relationship Between User and Product. In *Funology*, Mark A. Blythe, Kees Overbeeke, Andrew F. Monk, and Peter C. Wright (Eds.). Kluwer Academic Publishers, Norwell, MA, USA, 31–42. https://doi.org/10.1007/1-4020-2967-5_4
 - [19] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 203–212. <https://doi.org/10.1145/1753326.1753357>
 - [20] Kurtis Heimerl, Brian Gawalt, Kuang Chen, Tapan Parikh, and Björn Hartmann. 2012. CommunitySourcing: Engaging Local Crowds to Perform Expert Work via Physical Kiosks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1539–1548. <https://doi.org/10.1145/2207676.2208619>
 - [21] Hyeonsu B. Kang, Gabriel Amoako, Neil Sengupta, and Steven P. Dow. 2018. Paragon: An Online Gallery for Enhancing Design Feedback with Visual Examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 606:1–606:13. <https://doi.org/10.1145/3173574.3174180>
 - [22] Louise H. Kidder and Michelle Fine. 1987. Qualitative and quantitative methods: When stories converge. *New Directions for Program Evaluation* 1987, 35 (1987), 57–75. <https://doi.org/10.1002/ev.1459>
 - [23] Nigel King, Joanna Brooks, and Saloomah Tabari. 2018. Template Analysis in Business and Management Research. In *Qualitative Methodologies in Organization Studies: Volume II: Methods and Possibilities*, Malgorzata Ciesielska and Dariusz Jemielniak (Eds.). Springer, Cham, Switzerland, 179–206. https://doi.org/10.1007/978-3-319-65442-3_8
 - [24] Yuki Koyama. 2016. Computational Design Driven by Aesthetic Preference. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16 Adjunct)*. ACM, New York, NY, USA, 1–4. <https://doi.org/10.1145/2984751.2984781>
 - [25] Ranjitha Kumar, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R. Klemmer, and Jerry O. Talton. 2013. Webzeitgeist: Design Mining the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3083–3092. <https://doi.org/10.1145/2470654.2466420>
 - [26] Benjamin Lafreniere, Andrea Bunt, Matthew Lount, Filip Krynicki, and Michael A. Terry. 2011. AdaptableGIMP: Designing a Socially-adaptable Interface. In *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology (UIST '11 Adjunct)*. ACM, New York, NY, USA, 89–90. <https://doi.org/10.1145/2046396.2046437>
 - [27] Walter S. Lasecki, Juho Kim, Nick Rafter, Onkur Sen, Jeffrey P. Bigham, and Michael S. Bernstein. 2015. Apparition: Crowdsourced User Interfaces That Come to Life As You Sketch Them. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1925–1934. <https://doi.org/10.1145/2702123.2702565>

- [28] Talia Lavie and Noam Tractinsky. 2004. Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. *International Journal of Human-Computer Studies* 60, 3 (March 2004), 269–298. <https://doi.org/10.1016/j.ijhcs.2003.09.002>
- [29] Sang Won Lee, Rebecca Krosnick, Sun Young Park, Brandon Keelean, Sach Vaidya, Stephanie D. O’Keefe, and Walter S. Lasecki. 2018. Exploring Real-Time Collaboration in Crowd-Powered Systems Through a UI Design Tool. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 104:1–104:23. <https://doi.org/10.1145/3274373>
- [30] Sang Won Lee, Yujin Zhang, Isabelle Wong, Yiwei Yang, Stephanie D. O’Keefe, and Walter S. Lasecki. 2017. SketchExpress: Remixing Animations for More Effective Crowd-Powered Prototyping of Interactive Interfaces. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST ’17)*. ACM, New York, NY, USA, 817–828. <https://doi.org/10.1145/3126594.3126595>
- [31] Friedrich Leisch. 2008. Visualizing Cluster Analysis and Finite Mixture Models. In *Handbook of Data Visualization*, Chun-houh Chen, Wolfgang Karl Härdle, and Antony Unwin (Eds.). Springer, Berlin/Heidelberg, Germany, 561–587.
- [32] 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 (CSCW ’15)*. ACM, New York, NY, USA, 473–485. <https://doi.org/10.1145/2675133.2675283>
- [33] Paul A. Mabe and Stephen West. 1982. Validity of Self-evaluation of Ability: A Review and Meta-analysis. *Journal of Applied Psychology* 67, 3 (6 1982), 280–296. <https://doi.org/10.1037/0021-9010.67.3.280>
- [34] Wendy E. Mackay. 2004. The Interactive Thread: Exploring Methods for Multi-disciplinary Design. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS ’04)*. ACM, New York, NY, USA, 103–112. <https://doi.org/10.1145/1013115.1013131>
- [35] Joanna McGrenere and Gale Moore. 2000. Are We All in the Same “Bloat”? In *Proceedings of the Graphics Interface 2000 Conference*. 187–196.
- [36] John Michael Moore. 2007. *Design Exploration: Engaging A Larger User Population*. Texas A&M University, College Station, TX, USA. Doctoral Dissertation.
- [37] Michael J. Muller and Sarah Kuhn. 1993. Taxonomy Of PD Practices: A Brief Practitioner’s Guide. *Commun. ACM* 36, 6 (1993), 24–28. <https://doi.org/10.1145/153571.255960>
- [38] Michael Nebeling, Stefania Leone, and Moira C. Norrie. 2012. Crowdsourced Web Engineering and Design. In *Web Engineering*, Marco Brambilla, Takehiro Tokuda, and Robert Tolksdorf (Eds.). Vol. 7387. Springer, Berlin, Germany, 31–45. https://doi.org/10.1007/978-3-642-31753-8_3
- [39] Michael Nebeling, Maximilian Speicher, and Moira C. Norrie. 2013. CrowdAdapt: Enabling Crowdsourced Web Page Adaptation for Individual Viewing Conditions and Preferences. In *Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS ’13)*. ACM, New York, NY, USA, 23–32. <https://doi.org/10.1145/2494603.2480304>
- [40] Michael Nebeling, Maximilian Speicher, and Moira C. Norrie. 2013. CrowdStudy: General Toolkit for Crowdsourced Evaluation of Web Interfaces. In *Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS ’13)*. ACM, New York, NY, USA, 255–264. <https://doi.org/10.1145/2494603.2480303>
- [41] Jakob Nielsen. 1993. *Usability Engineering*. Academic Press, Cambridge, MA, USA.
- [42] Jakob Nielsen. 2000. Why You Only Need to Test with 5 Users. Jakob Nielsen’s Alertbox. <http://www.useit.com/alertbox/20000319.html>
- [43] Jonas Oppenlaender, Kristy Milland, Aku Visuri, Panos Ipeirotis, and Simo Hosio. 2020. Creativity on Paid Crowdsourcing Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20)*. ACM, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376677>
- [44] Judith Alice Redi, Tobias Hübelfeld, Pavel Korshunov, Filippo Mazza, Isabel Pova, and Christian Keimel. 2013. Crowdsourcing-based Multimedia Subjective Evaluations: A Case Study on Image Recognizability and Aesthetic Appeal. In *Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia (CrowdMM ’13)*. ACM, New York, NY, USA, 29–34. <https://doi.org/10.1145/2506364.2506368>
- [45] Christian Rohrer and John Boyd. 2004. The Rise of Intrusive Online Advertising and the Response of User Experience Research at Yahoo!. In *CHI ’04 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’04)*. ACM, New York, NY, USA, 1085–1086. <https://doi.org/10.1145/985921.985992>
- [46] Elizabeth B.-N. Sanders and Pieter Jan Stappers. 2008. Co-creation and the New Landscapes of Design. *CoDesign* 4, 1 (2008), 5–18. <https://doi.org/10.1080/15710880701875068>
- [47] Ugo Braga Sangiorgi, François Beuvs, and Jean Vanderdonckt. 2012. User Interface Design by Collaborative Sketching. In *Proceedings of the Designing Interactive Systems Conference (DIS ’12)*. ACM, New York, NY, USA, 378–387. <https://doi.org/10.1145/2317956.2318013>
- [48] Martin Schmettow. 2012. Sample Size in Usability Studies. *Commun. ACM* 55, 4 (April 2012), 64–70. <https://doi.org/10.1145/2133806.2133824>
- [49] Ben Shneiderman. 2000. Universal Usability. *Commun. ACM* 43, 5 (May 2000), 84–91. <https://doi.org/10.1145/332833.332843>
- [50] Ben Shneiderman. 2009. Creativity Support Tools: A Grand Challenge for HCI Researchers. In *Engineering the User Interface: From Research to Practice*, Miguel Redondo, Crescencio Bravo, and Manuel Ortega (Eds.). Springer, London, UK, 1–9. https://doi.org/10.1007/978-1-84800-136-7_1
- [51] Ben Shneiderman, Gerhard Fischer, Mary Czerwinski, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Tom Hewett, Pamela Jennings, Bill Kules, Kumiyo Nakakoji, Jay Nunamaker, Randy Pausch, Ted Selker, Elisabeth Sylvan, and Michael Terry. 2006. Creativity Support Tools: Report From a U.S. National Science Foundation Sponsored Workshop. *International Journal of Human-Computer Interaction* 20, 2 (2006), 61–77. https://doi.org/10.1207/s15327590ijhc2002_1
- [52] Stefan Thomke and Eric von Hippel. 2002. Customers as Innovators: A New Way to Create Value. *Harvard Business Review* 80, 4 (2002), 51–61.
- [53] Eric von Hippel. 2005. *Democratizing Innovation*. MIT Press, Cambridge, MA, USA.
- [54] Edwin B. Wilson. 1927. Probable Inference, the Law of Succession, and Statistical Inference. *J. Amer. Statist. Assoc.* 22, 158 (1927), 209–212. <https://doi.org/10.1080/01621459.1927.10502953>
- [55] Jane Wood and Denise Silver. 1989. *Joint Application Design*. Wiley, New York, NY, USA.

- [56] Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-experts. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1433–1444. <https://doi.org/10.1145/2531602.2531604>
- [57] Anbang Xu, Huaming Rao, Steven P. Dow, and Brian P. Bailey. 2015. A Classroom Study of Using Crowd Feedback in the Iterative Design Process. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1637–1648. <https://doi.org/10.1145/2675133.2675140>
- [58] Lixiu Yu, Aniket Kittur, and Robert E. Kraut. 2016. Encouraging “Outside-The-Box” Thinking in Crowd Innovation Through Identifying Domains of Expertise. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1214–1222. <https://doi.org/10.1145/2818048.2820025>
- [59] Lixiu Yu and Jeffrey V. Nickerson. 2011. Cooks or Cobblers?: Crowd Creativity Through Combination. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1393–1402. <https://doi.org/10.1145/1978942.1979147>

Received October 2019; revised November 2019; accepted December 2019