

Social Network, Web Forum, or Task Market? Comparing Different Crowd Genres for Design Feedback Exchange

Yu-Chun Grace Yen
Computer Science
University of Illinois
Urbana, IL
yyen4@illinois.edu

Steven P. Dow
Cognitive Science
UC San Diego
San Diego, CA
spdow@ucsd.edu

Elizabeth Gerber
Segal Design Institute
Northwestern University
Evanston, IL
egerber@northwestern.edu

Brian P. Bailey
Computer Science
University of Illinois
Urbana, IL
bpbailey@illinois.edu

ABSTRACT

Increasingly, designers seek feedback on their designs from crowd platforms such as social networks, Web forums, and paid task markets which demand different amounts of social capital, financial resources, and time. Yet it is unknown how the choice of crowd platform affects feedback generation. We conducted an online study where designers created initial designs and revised the designs based on crowd feedback. We measured the quantity, quality, and content of the feedback received at two iterations and from crowds driven by social status, enjoyment, and financial gain. Our results show, for example, that task markets yield more suggestions, online forums provide more process feedback, and social networks give the most suggestions without payment. We contribute an emergent framework for crowd feedback selection, opportunities for enhancing feedback services, and an experimental platform that researchers can adapt to reduce the burden of conducting online studies of design feedback.

Author Keywords

Feedback; critique; online crowds; crowdsourcing; design methods; creativity

ACM Classification Keywords

H.5.3 [Information Interface and Presentation]: Group and Organization Interfaces--Collaborative computing.

INTRODUCTION

Online crowd platforms offer unprecedented opportunities for designers to connect with potential users [20]. One reason to connect is to gather design feedback in domains such as Web, product, and interaction design. Online crowds enable designers to acquire feedback faster and from a more diverse user population than possible with face-to-face methods [33].

There are three main genres of crowd platforms from which designers can solicit feedback. One option is for a designer to post a feedback task to a market such as Amazon.com's

Mechanical Turk (MTurk) [1]. Providers responding to the task are typically motivated by the financial gain associated with the task (*financial* crowd). A second option is for the designer to solicit feedback from her own social network (e.g. via Facebook or e-mail) where providers are driven by their relation to the designer or an exchange of social capital [15, 31] (*social* crowd). A third option is to post the design to a Web forum (e.g. Reddit [3]) where providers have a shared interest in design or the topic and are motivated by enjoyment of the work (*enjoyment* crowd).

A central problem is a lack of empirical knowledge of how the choice of the genre of the crowd platform affects feedback generation. For instance, which crowd genre offers the most design suggestions, process insights, or best overall quality? This knowledge is critical for helping designers better allocate their financial resources, social capital, and time for acquiring the feedback desired. Researchers have compared financial and social crowds for task responses in other domains such as shopping advice [27], but it is unclear how these results apply to the generation of design feedback.

In this paper, we report results of an online study comparing the feedback generated by three genres of crowd platforms and for two iterations on a design. We recruited designers (N=22) creating visual designs for problems of their own choosing. Designers created initial designs and revised the designs based on the feedback received from MTurk (financial), designers' own social networks accessed via Facebook, Twitter, and email (social), and Reddit or other Web forums (enjoyment). We measured key attributes of the feedback including perceived quality, quantity, length, and valence; and categorized its content. We also measured the degree of revision between the design iterations and which iteration was most effective. Finally, we interviewed the designers to learn about their perceptions of the feedback.

Among many results, we found that a *financial* crowd provides feedback that is longer, is more positively valenced, and contains more design suggestions. An *enjoyment* crowd provides more process-oriented feedback and the most responses without payment. A *social* crowd provides more design suggestions than an enjoyment-driven crowd. All three crowds produced feedback of similar perceived quality. We also found that initial designs receive more responses and more questions about the goals of the work, whereas the revised designs receive more judgments of overall quality.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

DIS 2016, June 04–08, 2016, Brisbane, QLD, Australia

© 2016 ACM. ISBN 978-1-4503-4031-1/16/06 \$15.00

DOI: <http://dx.doi.org/10.1145/2901790.2901820>

Our work makes four contributions to the field of interactive systems design. First, we formulate our results into an emergent framework for recommending which crowd genre to solicit feedback from and at which iteration to maximize attributes of interest including quality, quantity, content category, and valence. Second, our results are the first to show that designers can apply crowd feedback to improve the perceived quality of their designs in a naturalistic context (i.e. outside a classroom exercise). Third, we share potential enhancements to existing crowd feedback services including increasing the transparency of the feedback exchange and being able to receive feedback from the same providers across design iterations. Finally, we built and deployed an experimental platform that researchers can adapt to reduce the burden of conducting online studies of design feedback. The platform offers the ability to host design iterations, generate sharable feedback links, collect the feedback, and evaluate the responses. Our framework could also be implemented in the platform to help designers more easily acquire feedback that suits their own individual needs.

RELATED WORK

We describe our contributions to the study of online crowds for feedback generation for interactive systems design. We also describe how our research is original relative to prior comparisons of different crowd genres and contrast our use of crowds for feedback generation to other uses for design.

Leveraging Online Crowds for Feedback

Designers can leverage online crowds to quickly access potential users and tighten evaluation cycles. Feedback can be solicited from crowds driven by financial, social, or enjoyment motivations. A financially-driven crowd can be accessed through research platforms such as Voyant [33], CrowdCrit [24], and Critiki [16], or commercial platforms such as FeedbackArmy [4] and UserZoom [5]. These systems implement various workflows and scaffolding techniques for generating the feedback, but they all leverage financial crowds. The advantages of leveraging a financial crowd include the ability to receive feedback on-demand, gain precise control over the amount of feedback received, and customize the evaluation prompts. A classroom study showed that designers can leverage the feedback from financial crowds to improve their designs in an iterative process [34]. The downside of using financial crowds is the cost. Although one instance of feedback generation is typically affordable (e.g. \$10 U.S. dollars) [33], generating feedback for many iterations could be cost prohibitive.

Rather than use a financial crowd, a designer could leverage their social network for feedback [15, 19]. One way to mediate feedback exchange with this type of crowd is to host a design on a Web platform such as RedPen.io [6] or UITests.com [7], and share the provided feedback link. A second approach is to post a design and its content to a social network site and generate feedback via the discussion model implemented on the site. The benefits of tapping a social

crowd include receiving feedback without financial cost, the feedback can be more diverse than face-to-face discussion [19, 20], and the social awareness between the designer and providers can aid interpretation. The disadvantages are that it conflates work with social life [20], can exhibit highly variable response rates [20, 30], and costs social capital [30]. Social capital is the resources available to an individual – such as the ability to gather design feedback – that can be extracted from his or her social network [10].

A third option is to post a design to an online discussion forum such as Reddit [3] or an online community such as Dribbble [2]. The benefit of participating in these forums is that a designer can reach an audience that shares a passion for design and/or the problem domain. The providers are motivated by enjoyment of the topic to give feedback. Though intrinsic motivation should lead to higher quality work outcomes [8], prior studies have found that the design feedback received from online communities can be of lower quality and quantity than designers expect [25, 32].

This corpus of prior work has each measured the utility of a specific crowd genre for generating feedback. In contrast, our work is the first to compare the feedback (e.g. quantity, quality, and content) generated by crowds driven by financial gain, social status, and enjoyment. Our work also examines how the iteration of the design affects the feedback. Our results will help designers better allocate their financial resources, social capital, and time for feedback acquisition.

Comparing Responses from Different Crowd Genres

Because technology is enabling access to different types of crowds, researchers have begun to compare the responses and behaviors of these crowds [18, 27, 28, 30]. For example, researchers have compared answers received for everyday questions between social and financial crowds and found that the responses were similar in content and quality [21, 30]. Morris et al. analyzed in-store fashion advice received from paid workers and the shopper's social network [27]. They found the feedback received from a financial crowd was perceived to be more honest and influential despite the lack of shared context and potential for privacy concerns. Another study compared the characteristics of daily living advice collected from a financial crowd and an online community. The financial crowd was found to provide more rapid and concise responses relative to the online community [18].

Our work shares the goal of comparing different crowds. However, our work compares three genres of crowds – financial, social, and enjoyment – and for the purpose of generating feedback on in-progress designs.

Engaging Crowds for Other Aspects of Design

Researchers have studied the use of online crowds in the design process for purposes other than feedback generation. For example, researchers have leveraged crowds to generate concept sketches [35], test task performance on prototypes [23], and synthesize a functional user interface in an end-to-

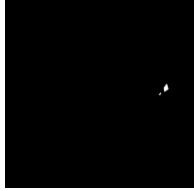



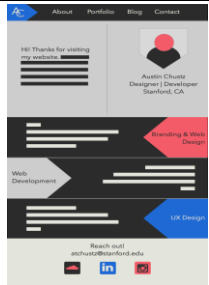

Category	Title	Initial	Revised	Description
Logo	Around the Coyote logo			This is a logo for an upcoming arts festival in Chicago called "Around the Coyote" - the festival has been on hard times and the original building named Coyote is currently being redeveloped to be condos. So the new logo divorces the old building to focus on the Flat Iron Arts Building and introduces a cartoon Coyote curled around the building protectively so that the name of the festival is retained and re-contextualized.
Poster	Exploratorium Bus Poster			This is an advertisement created for the Exploratorium's "After Dark" adult-nights.
Web	Austinchustz.me wireframe			This is intended to clarify the layout, colors, and general look and feel for my personal website. Eventually this will be the hub for my portfolio and blog as well as a way for people to contact me.

Figure 1. Examples of initial and revised designs created by designers in our study. The titles and descriptions listed were provided by the designers.

end process [29]. Dow et al. showed how students could benefit from crowd technologies at different design phases such as need-finding, ideation, and evaluation [14]. In [20], the authors further studied ten crowd-based design activities and found that these activities provided students with quick insights and feedback from authentic users.

Our work differs because it targets the use of crowds for the purpose of feedback generation. Feedback is a staple of all modern design processes. It helps designers learn about and improve a design, unblocks cognitive fixation, and fosters comparison between design alternatives [33], among other benefits. Crowd feedback can help designers iterate toward solutions that better satisfy the needs of potential users.

RESEARCH QUESTIONS

Our study compared three genres of crowds – financial, social, and enjoyment – for the generation of design feedback. We focused on three research questions.

RQ1: How do different crowd genres compare in terms of the quantity, quality, and content of the feedback generated?

RQ2: How does the design iteration (initial vs. revised) affect the feedback generated by the different crowds?

RQ3: What are designers' perceptions of getting feedback from the different crowd genres in the design process?

These questions are not exhaustive, but will help designers know how to better leverage different types of crowds during the design process and what the tradeoffs might be.

METHOD

To answer these questions, we conducted a full-factorial between-subjects online study. The factors were Iteration (Initial vs. Revised) and Genre (Financial vs. Social vs. Enjoyment), giving a 2x3 experimental design. Designers created initial designs, revised their designs based on the feedback received from the three crowds, and rated the quality of each piece of feedback. Our approach was primarily experimental because we wanted to gather and compare authentic design feedback and because we view online feedback generation as a relatively consistent and observable process. Interviews were conducted afterward.

Designers and Their Projects

Twenty-two designers participated in the study (ages 18–45, 10 female) and completed a background survey. 23% of the designers reported four or more years of professional experience in visual design, 40% reported one year or less, and the remaining 33% had between one and four years of experience. The average self-reported expertise was 3.1 on a 5-point scale (SD=0.89). For experience receiving feedback online, 60% reported having tried social networks such as Facebook or Twitter, 36% tried online design forums such as

Core77, Reddit and Dribble, and none of them had used paid task markets. Among all of the participants, sixteen lived in the U.S., two in India and four in the UK. They received \$70 for participation.

For the initial iteration, the designers either created a design for a visual design problem of their own choosing or used an in-progress design. All designs were revised based on the feedback received in the study. The projects included ten logos, eight Web pages, and four posters. Examples of the initial and revised designs for three of the projects are shown in Figure 1. Twenty of the designers reported that their design solutions were intended for a public audience.

Crowd Genres

Our study recruited feedback providers representing the different crowd genres using existing platforms. For a financially-driven crowd, we recruited providers from Amazon.com’s Mechanical Turk (MTurk). For the socially-driven crowd, designers posted their design to a social network site of their choosing and solicited feedback using email. Online forums identified by designers were used to approximate an enjoyment-driven crowd. If no forums were identified, we posted the designs to Reddit (the design_critiques subreddit). We chose Reddit as the default because it is an open discussion forum and has a large and active user base around design critique. Communities such as Dribbble or Behance were excluded because they require membership or reputation in order to receive feedback.

The main advantage of using existing platforms is ecological validity, as designers must choose between these types of platforms today. The disadvantage is the possibility of individual differences within each genre (e.g. not all providers from the task market are incentivized solely by financial gain). We return to this point in Discussion.

Experimental Platform

We built and deployed an online platform for hosting designs, collecting feedback from the crowds, and rating the feedback. A designer uploads an image of a design, describes its goals, and identifies it as initial or revised. See Figure 2.

Once the design is uploaded, our platform generates a sharable link for that design. When the link is selected, the feedback provider is presented with the form shown in Figure 3. On the form, the platform displays the image of the design along with the goals entered by the designer. If selected, the design image appears in a larger window. The form had three questions. First, the providers were asked to leave free-form feedback for improving the design. Second, the provider ranked three reasons for providing feedback corresponding to the assumed motives of each crowd genre in our study. Finally, the providers were asked to enter their gender, age, and self-rated design expertise. The first two questions were required, while the third set was optional.

Once the feedback was collected for a design, the designer returned to our platform and rated the perceived quality of

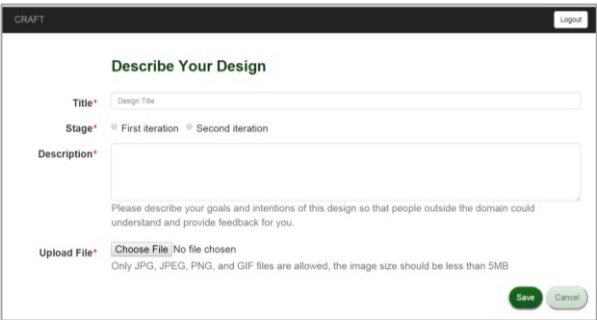


Figure 2. The user interface for a designer to upload a design, select the iteration, and describe its goals.

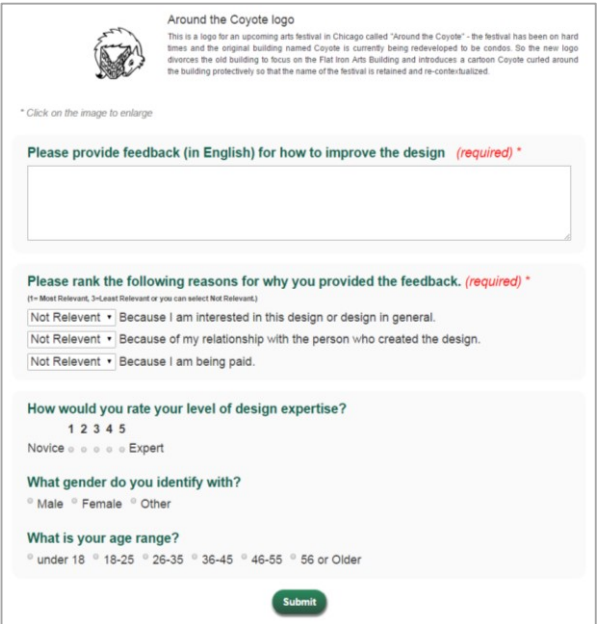


Figure 3. The user interface for a provider to leave feedback on a design and respond to brief survey questions.

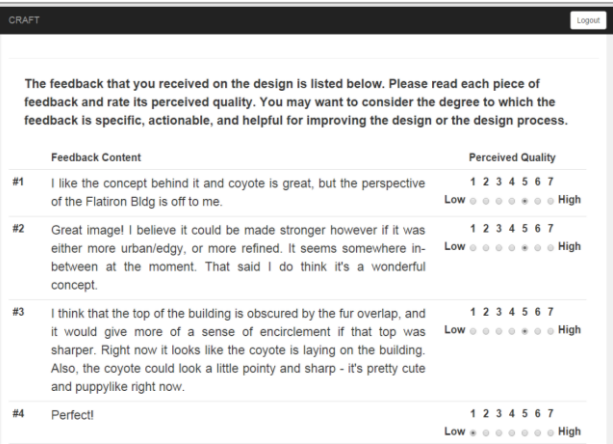


Figure 4. The user interface used by the designer for rating the perceived quality of the feedback responses.

each response on a scale of 1 (Low) to 7 (High). See Figure 4. The designers were free to apply their own criteria, but we suggested considering the degree to which each response was

Figure 5. After uploading a design, the designer was led through three steps: posting the feedback link to their social network sites, emailing the link to three people, and entering up to two online forums for receiving feedback. The research team then posted the design to the paid task market and the identified forums or the default.

specific, actionable, and helpful. Designers were blind to the crowd genre and identity of the providers to minimize bias based on prior experiences or pre-conceptions.

Procedure

After the informed consent process, the procedure consisted of two phases.

Phase I

In the first phase, the designer uploaded the initial design to our platform. The platform then displayed a three-step process for sharing the design. See Figure 5. In the first step, the designer posted his or her design to a social network site. The post contained the feedback link for the design and a short message. The suggested text was, ‘Hello all! I need your feedback on a design that I am currently working on. Please select the link to provide feedback. Thanks!’ The designer was allowed to modify the text if desired. Once complete, the designer entered where the design was posted into a field on the instruction page.

Second, the designer emailed the design and the suggested text to at least three people of their choosing. The step was included in case the designer had a limited social network. Last, the designer identified up to two online forums for sharing the design. If none were identified, the designer was informed that the design would be posted to Reddit.

To reduce the designer’s workload and remove the need for external accounts, the research team solicited the feedback from the platforms identified for the financial and enjoyment conditions. For the financial condition, we recruited 30 workers per design. Workers were paid \$0.50 per feedback task, which is a high wage relative to similar tasks in prior

extensive pilot testing, we found that users of online forums work (e.g. [24, 33]). For the enjoyment condition, through (especially Reddit) would not typically leave feedback through an external link. We therefore posted the design directly to the forum identified or Reddit (the default) and allowed the feedback to unfold on the platform. Designers were asked not to view the feedback. We waited at least six hours between design postings to reduce competition for responses. After two days, we copied the feedback from the forum to our platform to enable blind-to-condition ratings.

Once the feedback was ready to review, the designer was notified by email. The designer returned to our platform and rated the perceived quality of the feedback. See Figure 4. A sample of the feedback responses rated low (1) and high (7) by the designers is shown in Table 1.

Phase II

In the second phase, designers were given one week to revise their designs based on the feedback received from the first phase. Designers then repeated the three steps listed in phase one. However, the suggested text for the posting included a link to the first iteration of the design. For consistency, it was also included when requesting feedback on the platforms used for the paid and enjoyment conditions. At the end of this phase, we interviewed the designers. The questions covered the design (e.g. *what is the purpose of the design?*), perceptions of the feedback between iterations (e.g. *what differences did you notice between the two sets of feedback?*), and perceptions about the different crowd genres for design feedback (e.g. *between a paid platform, social network site, and online forum, which platform do you believe would provide the most helpful feedback? Which would you be most comfortable using and at which design iteration?*). The interviews were audio recorded and lasted 30 minutes. At the conclusion of the study, we deleted all postings made on online forums and informed designers that they could delete the postings made on their social network if desired.

Measures

Our study collected measures related to the changes between the design iterations, the perceived usefulness of the feedback, and the content of the feedback. All of these measures have been previously used for analyzing online design feedback [12, 34, 36].

Changes between Design Iterations

We recruited subjects to evaluate the initial and revised designs created by each designer. Subjects were shown a task screen displaying the designs side by side along with the goal entered by the designer. The placement of the designs was randomized. Subjects selected which of the two designs best achieved the stated goals. They also rated the degree of change between the designs on a seven-point scale ranging from *Almost the same* (1) to *Significantly different* (7). The experiment was implemented on MTurk. Ten workers were recruited to evaluate each set of designs (N=220, all unique IDs). We paid \$0.30 (US) per task.

	Initial Iteration	Revised Iteration
Financial	<p>[Rating=1] The web is a wonderful thing, have to create an organization details from the beginning. The web designed logo is attractive, need some more attractive one, have to create introduction page ,search engine , administration details etc.,</p> <p>[Rating=7] I think the contrast on the entire design is too high. I get what you're going for, but it makes it hard to read and just ends up looking washed out. Also, the teal font is hard to read on that background. Most of the text is actually. Besides those few things, the overall design is nice. Good fonts, good composition.</p>	<p>[Rating=1] I think that the design could be improved by using a more creative background instead of a plain white background. I also feel like instead of topic 1, topic 2, and topic 3, the topics should actually have a name to them. I also feel like the p cube need to be more creative.</p> <p>[Rating=7] What's confusing to me is that the "by type of school" graphic shows the percentage of kids who ARE vaccinated while the other two are showing the kids who have gotten waivers. People barely skim infographics, so I think it's important to be consistent. If I were just glancing at this, for instance, the title "Are Michigan Kids Vaccinated?" right next to "BY COUNTY" would make me think the darkest counties are the ones with the most vaccinated children, rather than the ones that have received the most waivers.</p>
Social	<p>[Rating=1] The design makes no sense to me.</p> <p>[Rating=7] Beautiful minimalism. The only direction I'd be curious to see is if you gave images a greater role throughout your portfolio. For example, including an image for each case study on page 2 would likely spark even more interest for me ... <i>(omitted)</i></p>	<p>[Rating=1] The design might be too simple, not very attracting. Maybe add a little more graphics?</p> <p>[Rating=7] It doesn't feel like the E or the R are actually in the design. Maybe just stick with the logogram of the name (though the y feels less interesting than the rest of the letters, and you can probably tighten up the kerning here) and either (1) make the design below based on the C... <i>(omitted)</i></p>
Enjoyment	<p>[Rating=1] I have no design background whatsoever but I'll add my 2 cents. I think it looks very aesthetically pleasing and I like the color scheme. However the first thing I thought was not skin care. Even having read your post I struggle to make the connection. Hope this helps.</p> <p>[Rating=7] Is there a reason you chose the vertical design if you move the lower third of information you would be able to fit it all on one page and have it be read able. as it is you have to scroll to really be able to digest all the information.</p>	<p>[Rating=1] "If we make the background box of the search results, it would look more like results and if we make the text after search icon lighter in color that would look better."</p> <p>[Rating=7] I like the layout of the design so far. However, the font choices need to be varied. It all looks to be the same font except for the lorem ipsum at the bottom. Vary the fonts (add a serif one perhaps for headers) and change your font families up (add bold, italics, or something) to give the page more interest... The color choices could be better as well. Right now...<i>(omitted)</i></p>

Table 1. Samples of the feedback generated by each crowd genre and at each design iteration. For the ratings (1=lowest perceived quality, 7=highest). Some content from the feedback was omitted for brevity.

Perceived Usefulness

Designers rated the usefulness of each feedback response on a 7-point Likert scale from 1 (Low) to 7 (High). Ratings were performed blind to the crowd genre that produced the feedback and identity of the provider.

Content Analysis

We measured the length and valence of the feedback and categorized its content. Valence captures the attraction (positive) or aversion (negative) that a person feels toward a feedback response. We recruited subjects from MTurk to label the valence of the feedback. Subjects from MTurk typically provide reliable labels for sentiment and content analysis [26]. We paid \$0.02 (US) per task. A subject was presented with one feedback response and labeled it from 1 (*strongly negative*) to 5 (*strongly positive*). Each response was labeled by two subjects (N=2876). If the response received identical labels, it was assigned that label. A majority (51%) of the responses received matching labels. The next 41% of the responses received labels that differed by one unit. In these cases, the disagreement was resolved by a member of the research team who cast the deciding vote. The remaining labels (8%) had larger disagreement and were resolved through discussion by the research team.

To categorize the feedback, we partitioned each response

into idea units and labeled the units based on a taxonomy of critique discourse described in [13]. An idea unit represents a coherent expression of thought [9]. Two coders with experience in HCI assigned the labels. Krippendorff's alpha, a measure of reliability between multiple raters and categories, was 0.85 on training samples, which represents very good agreement [22]. Therefore, the coders labeled the remaining units independently. We paid \$25 for each coder.

RESULTS

We report to what degree a designer improves his or her design based on the feedback from online crowds, how the crowd genre affects the quality, quantity, and content of the feedback generated at different design iterations, and designers' perceptions of the feedback received.

Revised Designs Improved

The revised designs were perceived to meet the stated goals more effectively than the initial designs. Out of the 220 votes, 148 were for the revised design while only 72 were for the initial design ($\chi^2(1, N=220)=26.3, p<0.001$). The revised design received a majority of the votes for fifteen of the twenty-two projects. The average rating of change between the two iterations was 4.41 (SD=0.83) on the seven-point scale. Taken together, these results show that a majority of the designers were able to leverage the feedback from the

Iteration	Financial	Social	Enjoyment	Total
Initial	30 (0)	2.83 (3.38)	4.08 (2.87)	36.6 (3.89)
Revised	30 (0)	1.65 (2.08)	2.04 (1.52)	33.7 (2.71)
Total	30 (0)	2.04 (2.55)	3.11 (2.62)	

Table 2. Mean and (SD) of the number of feedback responses by Genre and Iteration.

online crowds to make non-trivial changes to their designs, and these changes led to more effective designs. This corroborates results from a prior study of using crowd feedback in the classroom [34], but extends those results to a larger and more diverse collection of real-world designs.

A minority (seven) of the designers, despite making changes, were unable to create a revised design that was rated more effective than the initial design. This is not unexpected given that design can be viewed as a process of hill climbing, which can include temporary descents [11].

Initial Designs Received More Feedback

Summarized in Table 2, 811 responses were collected for the initial designs and 751 for the revised designs. The financial crowd was not included in this analysis since the number of responses was controlled by the research team. A repeated measures ANOVA with Genre (social vs. enjoyment) and Iteration (initial vs. revised) as factors showed that a designer receives nearly twice as many responses for an initial design ($u=3.5$) than for a revised design ($u=1.8$; $F(1,22)=25.5$; $p<0.001$). Genre had no effect and there were no interactions.

One explanation for this result is that the designers expended social capital on the feedback request for the initial design [30]. Providers on a social network site or online community may be unwilling to aid the designer again without reciprocity (this wasn't possible in our study since the exchange was anonymized) [31]. Providers may also be less willing to share feedback on the revised designs because they perceive less agency over the final solution.

Different Crowds Produced Feedback of Similar Value

Table 3 summarizes the ratings of perceived quality of the feedback responses. The ratings were analyzed using an ANOVA with Iteration (initial vs. revised) and Genre (financial vs. social vs. enjoyment) as factors.

Designers perceived the feedback for the initial designs ($\mu=4.38$) to have marginally more value than for the revised designs ($\mu=4.25$, $F(1,1561)=3.28$; $p=0.07$). One possibility is that the designers were less confident of the direction of the initial designs and therefore appreciated the feedback more. It may also have been easier for the providers to critique the initial designs since they were less polished.

Surprisingly, Genre did not affect the perceived value of the feedback. The distributions of the ratings were nearly identical in each condition. This means that designers can receive feedback of roughly similar quality from any of the crowds tested in our study.

Iteration	Financial	Social	Enjoyment	Total
Initial	4.35 (1.70)	4.98 (1.52)	4.25 (1.73)	4.38 (1.7)
Revised	4.26 (1.66)	4.74 (1.33)	4.07 (1.80)	4.25(1.67)
Total	4.30 (1.68)	4.72 (1.44)	4.18 (1.74)	

Table 3. Mean and (SD) of the perceived quality ratings by Genre and Iteration.

For the financial and social crowds, self-reported expertise did not correlate with perceived quality (Spearman's rank correlation coefficient was 0.03, n.s.). One explanation is that provides approach giving feedback based on the expectations associated with the platform. It also is possible that providers do not differentiate their expertise when giving design feedback online or may not accurately report it.

Content Analysis

Financial Crowd Produced Longer Feedback

For this analysis, given the high variability, we excluded responses that were two standard deviations away from the mean response length for each crowd genre. The final data set included 1493 responses (775 for initial and 718 for the revised designs). An ANOVA showed that Genre had a main effect on the length ($F(2,1493)=6.01$; $p=0.003$). Pairwise comparisons using Tukey's HSD showed that the length of feedback from the financial crowd ($\mu=61$ words) was longer than the feedback from the enjoyment crowd ($\mu=51$ words; $p=0.004$). The length of the feedback from the social crowd ($\mu=56$ words) did not differ from the others. No other effects were found. A financial crowd may have written more because they felt obligated to perform work commensurate with their payment. We also found a positive correlation (Pearson's $r=0.35$, $p<0.001$) between feedback length and its perceived usefulness, which echoes a similar finding in [36].

Financial Crowd Produced the Most Positive Feedback

Table 4 summarizes the valence ratings. An ANOVA revealed a main effect of Genre ($F(2, 1561)=6.17$; $p=0.002$). A post-hoc test using Tukey's HSD showed that the financial crowd wrote feedback that was more positive ($\mu=3.8$) than the social crowd ($\mu=3.5$; $p=0.001$) and the enjoyment crowd ($\mu=3.6$, $p=0.04$). The financial crowd may have written more positive feedback to ensure their responses would be accepted for payment. There were no other effects.

We tested the correlation between feedback valence and its perceived usefulness. In [36], the researchers found that design students rated positive feedback more useful. In our study, the Spearman's ρ indicated no relationship between the two variables ($\rho=0.01$; n.s.). One difference is that the designers in our study had professional experience and may therefore have learned to focus on the content rather than the valence of the feedback when judging its value.

Different Crowds Produce Different Categories of Content

We excluded the feedback from designs that did not receive responses from all three platforms for both iterations to enable fair comparison. The analysis therefore contained ten designs. For each iteration, we included up to the first ten

Iteration	Financial	Social	Enjoyment	Total
Initial	3.78 (1.10)	3.59 (0.94)	3.66 (1.37)	3.76 (1.13)
Revised	3.82 (1.04)	3.43 (1.17)	3.42 (1.80)	3.78 (1.09)
Total	3.80 (1.08)	3.52 (1.04)	3.58 (1.42)	

Table 4. Mean and (SD) of valence ratings by Genre and Iteration. Larger values are more positive.

responses from each genre to balance quantity. The final dataset contained 585 and 516 idea units for the initial and revised designs, respectively. For initial designs, the idea units were derived from 100, 42, and 39 responses from the paid, social, and enjoyment crowd, respectively and 100, 31, and 22 from the respective genres for the revised designs. Table 5 summarizes the idea units by content category.

To compare the patterns of interest, we performed z-tests for population proportions. We found that feedback on the initial designs contained more *investigations* (2.7%) than the feedback on the revised designs (0.9%, $z=2.13$, $p=0.03$). An *investigation* is when a provider questions specific aspects of the design or process. For example, “*Are any of these buildings IN Grand Rapids [a city in the state of Michigan in the U.S.]?*” Not surprisingly, designs with a less polished appearance invite more questions. From Table 5, the enjoyment crowd provided the most *investigations* while the financial crowd gave the fewest for both iterations. The financial crowd may have asked fewer questions because they did not expect further interaction with the designers.

Judgments appeared in the revised design (59.6%) more than in the initial design (53.8%, $z=1.9$, $p=0.05$). A *judgment* is when a provider assesses a design. For example, “*It’s a bit too complex. The final position of the design does not have enough symmetry.*” The more polished appearance of the revised design may have attracted more judgment. We also found that the financial crowd made more judgments for revised designs (62.6%) than initial designs (54.1%, $z=2.12$, $p=0.03$). In contrast, the social and enjoyment crowds generated similar numbers of judgments between iterations.

The enjoyment crowd provided more *process-oriented* feedback (16%) than both the social (6%, $z=3.61$, $p=0.003$)

and financial crowd (0.4%, $z=9.48$, $p<0.0001$) for both iterations. A *process-oriented* statement is when a provider references the process by which the design came about. For example, “*Websites aren’t slideshows. My advice? Study how websites are made, how they are built and structured.*” This difference could be due to the enjoyment crowd having more experience providing design feedback online. In contrast, the enjoyment crowd gave fewer *recommendations* (18%) than the financial crowd (33%, $z=3.65$, $p=0.0002$) for both initial and revised designs. The number of recommendations provided by the social crowd was in-between the others (24%). A recommendation is when a provider suggests a specific change to the design such as “*Let the text underneath the logo spell out its name.*”

Comparison is when a provider contrasted the design or design process with the prior iteration or an existing design. For example, “*I really like the minimal feel of your previous version.*” In our dataset, comparisons were only found for the revised designs. This may have been prompted by including a link to the initial design in the feedback request for the revised design. This observed effect may of interest for future research. We did not observe other patterns of research interest between the crowd genres or design iterations.

INTERVIEW RESULTS

Nineteen designers (out of 22) completed the interviews. Responses to the interview questions highlight the benefit and limitations of receiving feedback from online crowds. One benefit was being able to leverage crowd feedback to make evidence-based arguments during face-to-face design discussions: “*...with these feedback, I can finally convince my boss to remove the unnecessary background image*” [P24]. A second benefit was to acquire feedback that was perceived to be more open and honest than could be obtained from a face-to-face discussion: “*I think the platform gave me much more honest and diverse kind of viewpoints than I would get from my friends.*” [P19]. A third benefit was that the crowd feedback not only helped the designers to identify problems with the designs, but also to prioritize those problems: “*I see multiple people mention the same thing, and that to me, I think just not like a single person give me the*

Category	Initial Iteration				Revised Iteration			
	Financial	Social	Enjoyment	Total	Financial	Social	Enjoyment	Total
Judgment	54.1% (156)	54.8% (107)	50.9% (52)	53.8% (315)	62.6% (206)	58.9% (66)	48% (36)	59.6% (308)
Recommendation	35.7% (103)	24.1% (47)	21.5% (22)	29.4% (172)	29.1% (96)	25% (28)	13% (10)	25.9% (134)
Investigation	0.7% (2)	4.6% (9)	4.9% (5)	2.7% (16)	0	1.7% (2)	4% (3)	0.9% (5)
Interpretation	2.7% (8)	2.5% (5)	0.9% (1)	2.4% (14)	1.5% (5)	2.6% (3)	0	1.5% (8)
Brainstorming	3.1% (9)	3% (6)	2.9% (3)	3% (18)	3% (10)	2.6% (3)	0	2.5% (13)
Process	0	7.6% (15)	13.7% (14)	5% (29)	0.9% (3)	3.5% (4)	20% (15)	4.2% (22)
Comparison	0	0	0	0	1.2% (4)	0	5.3% (4)	1.5% (8)
Association	2.7% (8)	0.5% (1)	2.9% (3)	2% (12)	0.6% (2)	1.7% (2)	4% (3)	1.4% (7)
Identity-invoking	0	0	0	0	0	0	0	0
Total Idea Units	288	195	102	585	329	112	75	516

Table 5. Frequencies of the categories of idea units by Genre and Iteration.

same feedback, but many people think the similar way, and kind of carries more weight to me to consider it” [P7].

Designers also identified several limitations of interpreting crowd feedback. First, designers wanted to know more about the provider’s relevant background and experience: *“I wanted to listen to someone who is better than me, who I can trust rather than random people” [P2].* Similarly, another designer said: *“Well, there is a lot of feedback on the building but that’s what the building looks like, so there is not much I can do with that...I felt like some of the feedback is obvious people not necessarily in the audience that I design for, I might get more directly feedback from people who have already know the specific domain” [P25].* Crowd feedback systems should therefore consider soliciting relevant information from the providers and allowing the designers to cluster and filter the responses based on the collected data.

Second, the feedback was not always consistent with the iteration of the design. For instance, one designer stated, *“...I feel that in the revised iteration there were a lot more general feedback than specific suggestion, which I got a lot from my first design, and I don’t know if that’s necessarily because people thought that it was a completed work and they just wanted to give me general feedback” [P8].* Crowd feedback systems should therefore enable the designers to identify the stage of the design and the feedback desired (conceptual vs. specific suggestions) and make the background information and expectations available to the providers. Finally, the designers suggested better continuity with the providers for subsequent design iterations. *“It is important to have feedback for the initial design rather than the revised one. But it would be nice if you can find people reviewing the previous iteration to view the current one.” [P15]*

Designers also reported their perception of the feedback received from each crowd genre and design iteration. Nine of the designers preferred the feedback given to the initial design over the revised design: *“I think the initial one gives me broader senses just because the sort of range of different opinions, not necessarily that every opinion is useful in directly design but at least I have the awareness of how people would approach it. For the second set, I think I could just go outside and ask my friends who have design background, they’ll probably give me the similar feedback, so I don’t think it worth that much” [P3].* Four designers felt the feedback was comparable between the iterations.

For online forums, designers (n=9) expected to receive higher quality feedback because they believed those feedback providers would be more experienced in design:

“I would say the online forum that filled with designers could provide higher quality feedback because they can speak design language. And they hopefully have their users and have other’s opinion.” [P22]

“I feel like those platforms are filled with people who already are in design...I think the way they give feedback is the way they would want to get feedback.” [P25]

Other designers (n=4) were discouraged to use online forums due to prior experiences with variable quality, superficial comments, and harsh criticism:

“...I think there were some in-between feedback from design forums, their recommendations make no sense to me although I believe they know about design but just more novice than me...” [P23]

“For Reddit, I think they don’t really give you specific thing they don’t like or do like, they just either really like it or really don’t which isn’t really helpful for the designers.” [P22]

“I was always nervous about using Reddit because people there are harsh and also it doesn’t guarantee how much I would get.” [P18]

The perception of feedback received from a social network depends on one’s network composition. For example, eight designers expected to receive the least useful feedback from a social network site. One designer stated, *“I think things like Facebook or Twitter probably wouldn’t have high quality because not everybody on my FB I know connected with designer and have that kind of knowledge.”* In contrast, two designers trusted their network based on prior experience: *“I actually have a really awesome network of friends that I met previous in my school, so I would ask them in person or emailing them, you know, over social media, but it’s kind of people that I already know and have built relationship with. People who I admire and I trust their opinion. They wouldn’t dance around would give me very specific feedback” [P23].* For the financial crowd, designers did not report perceptions due to unfamiliarity with the task market used, MTurk.

DISCUSSION AND FUTURE WORK

A goal of our study was to help designers understand the differences in feedback received between crowd genres, and how these differences might be mediated by the iteration of the design. We distill our results into an emergent framework for selecting which crowd genre to solicit feedback from (the italicized text refers to the feedback attribute of interest):

Valence of the feedback. If a designer fears criticism, which can affect self-efficacy and receptivity to the feedback [17], it is best to solicit feedback from a financial crowd. Our results showed that this type of crowd produced the most positive-valenced feedback and was not sensitive to the design iteration. The other two crowds were equally critical.

Process-oriented feedback. From the content analysis, if a designer wants to receive insight or observations about the process for creating the design, she should solicit feedback from enjoyment crowds. If it is an initial design, the designer could also choose to get feedback from their social network.

Recommendations for improvement. If this category of content is most desired, then the designer should request feedback from a financial crowd regardless of the iteration. If a designer wants to avoid the cost, the next best choice is to use his or her social network.

Judgment. If a designer wants the design to be assessed holistically, then a financial crowd should be selected. This is particularly true for designs beyond the first iteration.

Number of responses. Our results show that a designer receives more responses from an enjoyment crowd than his social network regardless of the iteration. If the number of responses is insufficient, the designer could supplement the responses with feedback from a financial crowd.

Length of feedback. We found that a financial crowd produces the most elaborative feedback and this crowd is the best choice regardless of the iteration of the design.

Quality of feedback. Given the similarity of the quality ratings in our study, a designer can expect to receive similar quality feedback from any of the crowd genres tested. The decision is not affected by the iteration of the design.

Designers can utilize these recommendations to gather feedback that best suits their individual needs. However, it is currently up to the designer to know which attributes of the feedback are most needed or important for the design. The framework should also be followed with caution at this time since it was derived from a single study (see Limitations).

For future work, we want to make these recommendations accessible by implementing them in a platform such as the one built for our study or prior work (e.g. [33]). A designer could select the feedback attributes of interest and the platform would route the feedback request based on the selections and design iteration. An implementation would also enable deeper analysis of how designers learn to apply our recommendations and develop their own patterns of use. We also want to explore whether it would be possible to maximize multiple feedback attributes in parallel or how to choose the more appropriate genre of crowd when the recommendations conflict.

We analyzed the feedback by crowd genre. The genre was a proxy for what motivates providers to leave feedback. To test this approximation, we analyzed the rankings given by the providers for leaving feedback (see Experimental Platform). The results showed that 49% of the providers from MTurk ranked the payment as their top choice. Likewise, 66% of providers recruited from designers' social networks ranked their relation to the designer as the top reason for giving feedback. Rankings for providers from the online forums were not collected since they did not leave feedback through our platform. To check if the reported reasons affected our results, we re-clustered all of the data by the provider's top-ranked motivation and repeated the analyses. The pattern of results was nearly identical. This suggests that crowd platforms cultivate certain expectations of behavior which could influence how a provider approaches giving feedback.

LIMITATIONS

Our work attributes differences in the feedback between crowd genres to different motivations of the providers. Other factors could also contribute such as the domain experience,

design expertise, and demographics of the providers, and any expectations associated with the platforms. Future work is needed to test how these factors influence giving feedback online and to study the generalizability of our results for different instances of the crowd genres tested in this work.

Our study did not ask for or track the identity of the feedback providers due to privacy concerns. It was therefore not possible to test the overlap in the feedback providers between iterations. We suspect there was little overlap from the financial and enjoyment crowds given their large user bases, but there may have been overlap from the social crowds. An interesting question for future work is how design feedback from the same and new providers compares across iterations.

Given our research questions, it was critical for the designers to rate perceived quality blind to the source of the feedback. It was therefore not possible to separate the design revisions prompted by each crowd genre. A future experiment could have designers create designs and receive feedback from only one genre of crowd. This would allow the revisions to be compared between the crowds. Another limitation is that our work only considered two iterations of a design. Future work should therefore study how the feedback from different crowd genres is best applied for an entire design cycle. Our study included different design categories (e.g. logos, Web pages, and posters), but the sample size was too small to meaningfully compare the feedback between them. Finally, our study included designers ranging in experience from novice to early career stage. We look forward to testing the generalizability of our findings for designers with more domain expertise, different levels of experience with crowd feedback platforms, and different sizes of social networks.

CONCLUSION

Designers are increasingly accessing social networks, Web forums, and paid task markets to get feedback from potential users. But how does the choice of platform affect the design feedback received? To address this critical question, our work compared the quantity, quality, and content of feedback received from three genres of crowd platforms. Among many findings, we found that a paid task market provides feedback that contains more design suggestions and is longer and more positively-valenced. A Web forum provides more process-oriented feedback and the most responses without payment, while social networks provide more design suggestions. All three crowds produced feedback of similar perceived quality. We believe our results will enable and encourage designers to more effectively harness the immense potential of online technologies for feedback acquisition, leading to solutions that better satisfy the needs of potential users. The experimental platform can be accessed at <http://review-my-design.org/dis2016>. Interested readers can contact the first author for access to the source code.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under awards 1462693, 1530818, and 1213091.

REFERENCES

1. Amazon Mechanical Turk. <https://www.mturk.com/>
2. Dribbble. <https://dribbble.com/>
3. Reddit. https://www.reddit.com/r/design_critiques/
4. FeedbackArmy. <http://www.feedbackarmy.com>
5. UserZoom. <http://www.userzoom.com/>
6. Red Pen. <https://redpen.io/>
7. UI Test. <http://www.uitests.com/>
8. Teresa M. Amabile. 1985. Motivation and creativity: effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology*, 48, 2: 393-399. <http://dx.doi.org/10.1037/0022-3514.48.2.393>
9. Natasha Artemeva and Susan Logie. 2002. Introducing engineering students to intellectual teamwork: The teaching and practice of peer feedback in the professional communication classroom. *Language and Learning across the Disciplines*, 6, 1: 62-85.
10. Moira Burke, Robert Kraut and Cameron Marlow. 2011. Social capital on facebook: differentiating uses and users. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 571-580. <http://dx.doi.org/10.1145/1978942.1979023>
11. Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc.
12. Cristian Danescu-Niculescu-Mizil, Gueorgi Kossinets, Jon Kleinberg and Lillian Lee. 2009. How opinions are received by online communities: a case study on amazon.com helpfulness votes. *Proceedings of the International Conference on World Wide Web*, 141-150. <http://dx.doi.org/10.1145/1526709.1526729>
13. Deanna P. Dannels and Kelly Norris Martin. 2008. Critiquing Critiques: A Genre Analysis of Feedback Across Novice to Expert Design Studios. *Journal of Business and Technical Communication*, 22, 2: 135-159. <http://dx.doi.org/10.1177/1050651907311923>
14. Steven P. Dow, Elizabeth Gerber and Audris Wong. 2013. A Pilot Study of Using Crowds in the Classroom. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 227-236. <http://dx.doi.org/10.1145/2470654.2470686>
15. Rebecca Gray, Nicole B. Ellison, Jessica Vitak and Cliff Lampe. 2013. Who wants to know?: question-asking and answering practices among facebook users. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1213-1224. <http://dx.doi.org/10.1145/2441776.2441913>
16. Michael D. Greenberg, Matthew W. Easterday and Elizabeth M. Gerber. 2015. Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers. *Proceedings of the ACM Conference on Creativity & Cognition*, 235-244. <http://dx.doi.org/10.1145/2757226.2757249>
17. Emily Harburg, Julie Hui, Michael Greenberg and Elizabeth M. Gerber. 2015. Understanding the Effects of Crowdfunding on Entrepreneurial Self-Efficacy. *Proceedings of the ACM Conference on Computer Supported Cooperative Work* 3-16. <http://dx.doi.org/10.1145/2675133.2675142>
18. Hwajung Hong, Eric Gilbert, Gregory D. Abowd and Rosa I. Arriaga. 2015. In-group Questions and Out-group Answers: Crowdsourcing Daily Living Advice for Individuals with Autism. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 777-786. <http://dx.doi.org/10.1145/2702123.2702402>
19. Julie Hui, Amos Glenn, Rachel Jue, Elizabeth Gerber and Steven Dow. 2015. Using Anonymity and Communal Efforts to Improve Quality of Crowdsourced Feedback. *Proceedings of Human Computation and Crowdsourcing*.
20. Julie S. Hui, Elizabeth M. Gerber and Steven P. Dow. 2014. Crowd-based design activities: helping students connect with users online. *Proceedings of the ACM Conference on Designing Interactive Systems*, 875-884. <http://dx.doi.org/10.1145/2598510.2598538>
21. Jin-Woo Jeong, Meredith Ringel Morris, Jaime Teevan and Daniel Liebling. 2013. A Crowd-Powered Socially Embedded Search Engine. *The International Conference on Weblogs and Social Media*,
22. David E. Kanouse and L. Reid Hanson. 1972. Negativity in Evaluations. *General Learning Press*, Morristown, NJ, 47-62
23. Steven Komarov, Katharina Reinecke and Krzysztof Z. Gajos. 2013. Crowdsourcing performance evaluations of user interfaces. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 207-216. <http://dx.doi.org/10.1145/2470654.2470684>
24. 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. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 473-485. <http://dx.doi.org/10.1145/2675133.2675283>
25. Jennifer Marlow and Laura Dabbish. 2014. From rookie to all-star: professional development in a graphic design social networking site. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 922-933. <http://dx.doi.org/10.1145/2531602.2531651>

26. Tanushree Mitra, C.J. Hutto and Eric Gilbert. 2015. Comparing Person- and Process-centric Strategies for Obtaining Quality Data on Amazon Mechanical Turk. *Proceedings of the the ACM Conference on Human Factors in Computing Systems*, 1345-1354. <http://dx.doi.org/10.1145/2702123.2702553>
27. Meredith Ringel Morris, Kori Inkpen and Gina Venolia. 2014. Remote shopping advice: enhancing in-store shopping with social technologies. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 662-673. <http://dx.doi.org/10.1145/2531602.2531707>
28. Anne Oeldorf-Hirsch, Brent Hecht, Meredith Ringel Morris, Jaime Teevan and Darren Gergle. 2014. To search or to ask: the routing of information needs between traditional search engines and social networks. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 16-27. <http://dx.doi.org/10.1145/2531602.2531706>
29. Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine and Michael S. Bernstein. 2014. Expert crowdsourcing with flash teams. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 75-85. <http://dx.doi.org/10.1145/2642918.2647409>
30. Jeffrey M. Rzeszotarski and Meredith Ringel Morris. 2014. Estimating the social costs of friendsourcing. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 2735-2744. <http://dx.doi.org/10.1145/2556288.2557181>
31. Edwina S Uehara. 1995. Reciprocity reconsidered: Gouldner's "moral norm of reciprocity" and social support. *Journal of Social and Personal Relationship*, 12, 4: 483-502. <http://dx.doi.org/10.1177/0265407595124001>
32. Anbang Xu and Brian Bailey. 2012. What do you think?: a case study of benefit, expectation, and interaction in a large online critique community. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 295-304. <http://dx.doi.org/10.1145/2145204.2145252>
33. Anbang Xu, Shih-Wen Huang and Brian Bailey. 2014. Voyant: generating structured feedback on visual designs using a crowd of non-experts. *Proceedings of the ACM Conference on Computer Supported Cooperative Work* 1433-1444. <http://dx.doi.org/10.1145/2531602.2531604>
34. Anbang Xu, Huaming Rao, Steven P. Dow and Brian P. Bailey. 2015. A Classroom Study of Using Crowd Feedback in the Iterative Design Process. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1637-1648. <http://dx.doi.org/10.1145/2675133.2675140>
35. Lixiu Yu and Jeffrey V. Nickerson. 2011. Cooks or cobblers?: crowd creativity through combination. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1393-1402. <http://dx.doi.org/10.1145/1978942.1979147>
36. Alvin Yuan, Kurt Luther, Markus Krause, Sophie Vennix, Steven P. Dow and Björn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1003-1015. <http://doi.acm.org/10.1145/2818048.2819953>