A Crowdsourcing Model for Receiving Design Critique

Abstract
Designers in many domains are increasingly turning to online communities to receive critiques of early design ideas. However, members of these communities may not contribute an effective critique due to limited skills, motivation, or time, and therefore many critiques may not go beyond "I (don't) like it". We propose a new approach for designers to receive online critique. Our approach is novel because it adopts a theoretical framework for effective critique and implements the framework on a popular crowdsourcing platform. Preliminary results show that our approach allows designers to acquire quality critiques in a timely manner that compare favorably with critiques produced from a well-known online community.

Keywords
Critique, Design, Crowdsourcing

ACM Classification Keywords
H5.3. Information Interfaces and Presentation: Group and Organization Interfaces.

General Terms
Design, Human Factors, Algorithms
Introduction
Posting designs to online communities is a convenient way for designers to obtain critiques of their work. However, there is no guarantee of the quantity or quality of critiques received. Also, online communities allow users to comment on designs independently and in a free-form manner, which is inconsistent with the theory of critique in art and design education [5, 12]. Consequently, many online critiques do not go beyond simplistic statements of “I like it” or “I don’t like it”, which offers little benefit to the designer (see Figure 1).

Research in art and design education has investigated the practice of critique for several decades [1, 4, 5, 6, 11, 12] and has shown that structuring the critique process can improve critique quality [4, 5, 11, 12]. In this paper, we show how to apply a known theoretical framework for critique within online practice in order to improve the quality of critiques received online.

The main idea of our approach is to convert the critique process into a series of sub-tasks using the sequential critiquing framework. These sub-tasks are submitted to a crowdsourced platform and the results of the individual efforts are selectively aggregated to form a complete critique. Our approach provides a low-cost way of receiving quality critiques quickly, and the approach could be implemented within many existing design tools such as Photoshop or Visio.

In the following sections, we introduce the sequential critiquing framework from research in art and design education and illustrate how to apply the framework to an online critique process. Then, we present a scenario of receiving a critique on the logo of the CHI conference 2011 via our approach. We conclude with a discussion of the results and future study.

Sequential Process for Critique
An important finding from art and design education is that the quality of a critique is preferred over quantity [1, 6]. Feldman [4, 5] developed a well-known sequential approach to promote critique quality. It is a four-step process including description, analysis, interpretation, and judgment. The main idea of this framework is to delay judgment and convert the critique process into sub-tasks. Teo and Chai [11] followed the Feldman’s idea and applied the framework in design studios. The simplified framework has four sub-activities (see Figure 2):

1) Describe the intent of the design and its audience.
2) Evaluate the strengths of the design.

Figure 1. The logo of the CHI 2011 conference. We submitted this logo to an active online critique community¹ and received two critiques. One was “I like the layout and color”, and the other was “the tagline in the logo is difficult to read”. These critiques only have a few sentences, lack in-depth analysis of the design, and offer little insight for consideration. In contrast, we receive a more comprehensive critique on this logo via the approach we implemented. (see Testing Our Approach for details).

¹ DeviantArt (http://www.deviantart.com) is ranked the first among art and design communities [8].
3) Evaluate the weaknesses of the design.

4) Summarize the statements in the preceding steps.

The sequential framework not only ensures the quality of critique but also divides the critique process into a series of sub-tasks, shedding light on mechanisms of aggregating individual efforts to make a complete critique.

![Figure 2. The sequential critiquing framework.](image)

**Crowdsourcing Model for Online Critique**

It is difficult to directly apply the above framework in the context of online critique communities. Although this framework can convert the critique process into sub-tasks, requesting community members to follow specific instructions and work on the sub-tasks iteratively would dramatically increase the micro-costs of contributing critiques, which would affect participation [2]. Also, implementing the framework would also require large changes to the underlying system software driving the community.

Fortunately, iterative tasks defined by the framework could be efficiently performed in a for-pay human computation system like Amazon Mechanical Turk (MTurk) [9, 10]. We apply the above framework to MTurk by converting the critique process into a series of sub-tasks and designing various voting schemes to ensure the responses meet the requirements. The framework is implemented in JavaScript and the API provided by Turkit [9].

*Description*

This step is used to gather information and help later workers (Turkers) make in-depth evaluation in the next steps. In order to avoid the increased cost of requesting critique, we do not require a designer to elaborate his/her own design intentions. Our approach instructs workers to accomplish this task, since a designer's intentions can be described by others [11].

Specifically, workers are required to view the design and answer the following questions: 1) What do you believe are the designer's intentions? 2) What is the target audience of the design? After the first worker answers these questions, subsequent workers will be instructed to improve the initial description by either revising the writing of the description or adding new ideas to the description. Since workers may have different styles and diverse aesthetic experiences, we adopt a voting scheme to prevent the introduction of errors [9]. Every improvement made by a worker will be voted on by at least two other workers. If both of the workers vote for the improvement, it will be kept for the next iteration; if both of them vote against the...
improvement, it will be rejected; and if the two votes disagree, then a tie-breaking vote will be cast by another worker to decide whether the improvement will be kept.

**Strength**

In the step, workers are first required to view both the design and its description generated by previous workers. Then, workers are required to point out which element in the design supports the designer’s intention and why this element reinforces the design concept. In other words, workers are instructed to keep a positive attitude, find the strengths of the design, and relate the strengths to the designer’s intent. We create five separate assignments in this step and each assignment requires a worker to provide at least one strength of the design. As a result, the five assignments will produce at least five strengths of the design.

**Weakness**

Similar to the prior step, this step “weakness” first asks workers to view the design and the corresponding description. Then the workers are instructed to identify elements in the design that do not effectively realize or support the designer’s intent. The workers are required to reveal the weaknesses of the design and offer corresponding suggestions. The steps “weakness” and “strength” can be performed in parallel.

**Summarize**

The final step consists of two stages: select and aggregate. In the first stage, workers view the design and its corresponding description, and vote on each strength and weakness to indicate whether they agree with the statement. Multiple independent agreements suggest that the statement is reasonable. Our approach aggregates the votes and ranks the statements. The top three strengths and weaknesses are selected for the aggregation stage.

During aggregation, workers are asked to summarize the strengths into one paragraph and the weaknesses into another. Multiple versions of the summaries are generated, but only one summary of the strengths and weaknesses is selected based on subsequent workers’ votes. The final version of the critique consists of three paragraphs: description, strength, and weakness.

**Testing Our Approach**

The goal of our initial study was to test the viability of the proposed approach to receive critiques online. As the sequential framework was originally developed for critiquing visual designs, we use the logo of the CHI 2011 conference, a representative visual design, to test our approach. The critique received is as follows:

**Description.** "The logo has text in dark blue-green, faded blue-green and yellow. The dark color showing “CHI” is the first recognized and that shows the importance of the organization. The faded blue-green “20” of 2011 overlaps the dark blue-green C which changes the CHI into a unique logo to represent just the conference. The "11" of 2011 is smaller, yellow and in a different font compared to the same font being used for the rest of the logo. The entire logo represents integration and differences in color and boldness make the more important information easier to remember - "CHI" and "11". The target audiences are likely to be those who know what CHI is, such as members of CHI or organizations and people that deal with CHI because they work in a similar field. The logo promotes the conference to the target audience in the hopes they will attend."
**Strength.** “The logo is very popping. The 20HI stood out first, after it engaged me it I saw the CHI and then the 11, the next thing i looked at was the Vancouver, BC. It flows well and is intriguing to the eyes. Another thing that caught my eye is that the 20HI kinda plays a trick with your eyes make me think 2011. I think it looks great the way it is.”

**Weakness.** “While the logo is overall well done, better clarity could be achieved through the following adaptations: First, the light blue is a weak color in contrast to the teal-- changing all of the light blue text to pale orange would be a nice substitute. I would remove the italics on the "11," increase the font size, and change it to a brown or tan. Also, it is difficult to read Vancouver, BC May 7-12, 2011. A thin outline around the text would allow the date and location to stand out more. Lastly, the overall flow of the logo is nice but I do not think the visual relationship between "20" and "11" is clear. The yellow "11" seems somewhat out of place, and it might be easier to tell that it's intended to be part of 2011 with different placement.”

Interestingly, we found that two workers in the “strength” and “weakness” steps copied content directly from the conference website (www.chi2011.org) and submitted it as a statement of strength or weakness. The voting scheme in the later step filtered these statements successfully.

Our approach cost $1.41 US dollars and recruited 32 workers to accomplish the critique. It took about 7 hours (1.5 hours for the “description” step, 3 hours for the “strengths” and “weakness” steps, and 2.5 hours for the “summarize” step) to receive the critique. The completion time could be considerably reduced by increasing payment for the sub-tasks [7].

The presented critique systematically analyzes the logo design and offers constructive suggestions for change. Compared with the critiques obtained from the online community (see Figure 1), the quality of the critique is significantly improved. We also tested the approach on several visual designs such as posters and the results suggest that the framework is robust and reliable.

**Discussion**

Participating in online critique communities can yield benefits not captured in our approach. For example, long-term participation in a critique community can foster professional and social connections and establish reputation. Also, the act of writing a critique or the exposure to critiques written by others promotes learning within the discipline [6]. However, our approach provides equal opportunity for everyone to receive critique and is independent of an individual’s social network. More importantly, designers can leverage our approach to receive a comprehensive and well-organized critique quickly rather than a series of statements that may be overly simplistic or redundant.

Although our approach, in general, is able to help designers acquire critiques on various kinds of visual designs such as graphic design and web design, we notice that certain workers may not have sufficient knowledge or expertise to accomplish a task in the framework. One possible solution is to develop verifiable questions and qualification tasks to select appropriate workers and improve the likelihood of high-quality answers [7].

In addition, we found that the “summarize” step may produce summaries of various length without changing the meaning of the original statements. Bernstein et al.
[3] proposed the Find-Fix-Verify method, which can call on Mturk workers to shorten the text without altering its meaning. Our framework can adopt this method to allow the designer to choose how concise the critique should be balanced against additional time and cost.

There are at least three compelling directions for future work. First, it would be interesting to perform an empirical comparison among critiques received from online communities, critiques received from experts, and critiques received from our approach. Performance will be measured by counting the number of unique insights and the quality of the insights received in each condition. Additionally, a post-session questionnaire will be used to assess the designers’ satisfaction. Second, the promising results suggest a greater role of online critique in the design process. We plan to integrate our approach into an existing design tool such as Photoshop and conduct a longitudinal study to understand how the use of the tool affects the design process. Third, we present our model as an instance of harnessing the wisdom of the crowd to facilitate creative design activities and we believe crowdsourcing could play a more general role in the creative design process.

Acknowledgements
We gratefully acknowledge the helpful comments of the anonymous reviewers. This work was supported in part by the National Science Foundation under awards no. IIS 06-13806 and 06-43512.

Citations