Decipher: An Interactive Visualization Tool for Interpreting Unstructured Design Feedback from Multiple Providers

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

Joy O. Kim
Creative Intelligence Lab
Adobe Research
San Francisco, CA
joykim@adobe.com

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

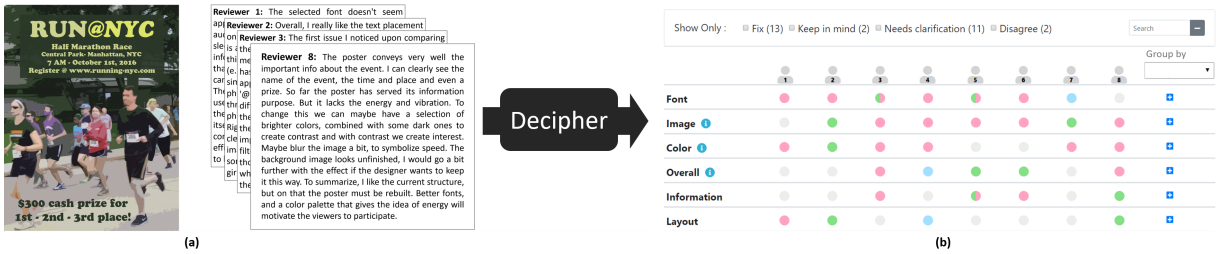


Figure 1: In (a), a designer has created a preliminary design and received feedback from multiple providers. In (b), the designer has imported the feedback into Decipher to visualize the topic and sentiment structure of the feedback. The designer can identify the strengths and weaknesses of different aspects of the design (row-wise comparison) and compare opinions between providers (column-wise comparison) without having to read the content itself. The designer can also interact with the content (e.g., to mark statements to incorporate in a revision or that need further clarification).

ABSTRACT

Feedback from diverse audiences can vary in focus, differ in structure, and contradict each other, making it hard to interpret and act on. While prior work has explored generating quality feedback, our work helps a designer *interpret* that feedback. Through a formative study with professional designers (N=10), we discovered that the interpretation process includes categorizing feedback, identifying valuable feedback, and prioritizing which feedback to incorporate in a revision. We also found that designers leverage feedback topic and sentiment, and the status of the provider to aid interpretation. Based on the findings, we created a new tool (Decipher) that enables designers to visualize and navigate a collection of feedback using its topic and sentiment structure. In a preliminary evaluation (N=20), we found that Decipher helped users feel less overwhelmed during feedback interpretation tasks and better attend to critical issues and conflicting opinions compared to using a typical document-editing tool.

Author Keywords

feedback; creativity; sense-making; creativity support tools.

CCS Concepts

•**Human-centered computing** → **Interactive systems and tools;** *Interface design prototyping;*

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376380

INTRODUCTION

In creative work such as design, writing, and music, gathering feedback from a diverse audience is a critical part of the creative process; people with different backgrounds and expertise may perceive the same creative work in different ways. To support this, HCI researchers have developed many new tools and interventions for generating constructive feedback from diverse providers [16, 17, 25, 29]. As a result, high-quality feedback is increasingly easier to create and discover.

However, interpreting and acting on a large collection of feedback received from providers with different backgrounds and opinions remains difficult. Prior work, for example, has shown that novices often fail to improve their work despite making moderate revisions after receiving high-quality feedback [33], even in communities that specialize in feedback exchange [10]. Feedback from diverse audiences may contain contradictions, focus on different topics, and vary widely in structure, making it hard to find emerging patterns, reconcile conflicting ideas, and prioritize revisions.

In this work, we study, design, and implement techniques to mitigate the difficulty of interpreting multiple pieces of feedback in the domain of graphic design. First, through a formative study (N=10), we found that even professional designers find the task of organizing and integrating multiple pieces of feedback into a single project overwhelming and that the feedback is often received in free-form formats (such as through email, PDF annotations, or typed notes taken in face-to-face meetings). We also discovered that experts employ three common strategies to interpret multiple pieces of design feedback: they *identify* the valuable feedback statements, *categorize* the statements, and *prioritize* the issues they find. When

performing these strategies, designers frequently reference the perceived sentiment and topic of the feedback statements, and the status of the provider who wrote those statements.

To demonstrate how the strategies discovered in the formative study could be embodied in a tool, we created Decipher, an interactive visualization tool that partitions written feedback into individual statements and visualizes the topic and sentiment structure of those statements across feedback providers (Figure 1). The visualization is generated using metadata collected through a labeling process. The tool also allows the designer to annotate emergent interpretations of the content so he or she can search and filter those annotations when preparing revision plans. Like other research that aggregates design feedback [33, 41], we leverage visualization to surface patterns in the data and aid sensemaking [7]. However, rather than structuring the feedback generation process to produce the visualizations (e.g., through micro-tasks [41] or by providing pre-authored statements [33]), our work focuses on visualizing the unstructured feedback designers already often receive from multiple sources (e.g., from clients, end users, and colleagues).

To evaluate how a tool like Decipher affects novice interpretation of design feedback, we conducted an exploratory evaluation ($N = 20$) where the participants used both Decipher and a typical document-editing tool (i.e., Google Docs) to interpret a collection of feedback given to two event marketing flyers. We focus on novices in this first evaluation because we felt novices would have the most to gain from a tool that scaffolds feedback interpretation. Our results show that Decipher helped novices feel less overwhelmed while engaging with the feedback and helped novices better identify critical issues, attend to strengths and weaknesses, and locate specific suggestions in the feedback relative to using the baseline tool. Novices also preferred the use of Decipher over the baseline tool for performing the feedback interpretation tasks.

In this work, we focus on graphic design as an example of a domain where feedback plays a prominent role in the creative process, but imagine the contributions of this work can be generalized to other domains where an individual receives multiple pieces of unstructured feedback for iterating on their creative work, such as in writing, music, or research. We hope to show that the HCI community, as designers of platforms and tools for creative work, can not only facilitate feedback generation but can (and should) create new tools and mechanisms to help creators "decipher" critical insights across feedback from multiple providers.

RELATED WORK

This paper builds on and significantly extends prior work addressing the difficulty of generating, interpreting, and acting on feedback received from diverse audiences.

Supporting feedback interpretation

Receiving good feedback early in the process of creative work is essential to the success of iterative design processes [19]. However, for feedback to be effective, it requires the recipient to interpret, learn, and act on it [24, 39]. Prior work has shown that novices are less likely than experts to spot opportunities for improving their creative work when reviewing feedback from multiple providers [15]. One reason is that novices often lack the requisite domain knowledge for analyzing information sources and translating the feedback into concrete ideas for revising an in-progress solution [6, 21, 38]. This problem will only become more prevalent as people are increasingly able to gather diverse feedback through crowd-based platforms [43].

The HCI community has generally approached these barriers to feedback implementation through two threads of research. One thread has sought to improve feedback quality through interventions such as rubrics [29, 45], examples [25], directed questions [9, 17, 41], templates [11, 20], and pre-authored statements [33]. The other thread has sought to affect how a recipient engages with the feedback received, such as by performing a reflection [4, 44] or coping [40] activity regarding the feedback, or writing an action plan [23]. However, research in both threads has found that creative outcomes may not always improve even when supported with feedback that is considered high-quality [10, 33, 44]. This may due to the fact that the feedback contains differing and possibly conflicting perspectives of a work, as this can increase the cognitive demands of the interpretation task [35]. Making sense of feedback and integrating feedback into revisions typically requires adding structure to the feedback content [14]. Our work extends prior research by identifying the criteria that designers employ to structure feedback and contributing a tool for visualizing and interactively exploring that structure.

Visualizing feedback and other data

Information visualization is the process of representing large data sets in a visual and meaningful way so that a user can better identify patterns, communicate messages, and reason about problems [7]. For example, prior work found that people are more likely to agree and respect others' opinions when they navigated the associated comments via a graphical visualization compared to a textual list view [13]. Visualizations can also facilitate consensus-building for groups [32] and decision-making for individuals while solving programming problems [31]. We propose that representing design feedback in a visualization can similarly help feedback recipients better discover and reason about the patterns present in written feedback.

A key challenge is that design feedback is often ill-structured and highly individualized to the particular project and designer. Prior research has addressed the challenge of aggregating design feedback by structuring the feedback at the time of creation (for example, by asking the providers to write feedback in response to directed questions [9, 41] or select from a discrete set of feedback statements [33]). However, these approaches limit the scope of expressible opinions to those defined by the system creator. In addition, the primary goal of such systems is to convert many novice opinions to a functionally expert one through aggregation and summarization. In contrast, we focus on visualizing equally important and unstructured feedback received from multiple providers (e.g., clients and end users) to facilitate opinion comparisons and pattern finding rather than aggregating micro-task outcomes.

In broader literature on text visualization, there are approaches that extract and visualize attributes such as topic, sentiment, and term frequencies to help users explore text data [22, 30,

42]. For example, Opinion Observer visualizes the sentiment and topic structure within online reviews to foster comparisons between products referenced in those reviews [30]. RevMiner extracts and presents the attribute-value pairs from restaurant reviews to facilitate search and sensemaking of those reviews [22]. Decipher shares the goal of visualizing the topic and opinion structure within a collection of feedback, but additionally structures the feedback by the providers who wrote that feedback. Our tool also supports interactions for users to capture his or her intentions for acting on statements within the feedback. We also contribute empirical results of how different representations of design feedback affect sensemaking.

FORMATIVE STUDY OF FEEDBACK INTERPRETATION

To understand the strategies that experts use to interpret and act on feedback received from multiple providers for a design, we conducted interviews with ten professional designers (six female) recruited through email distribution lists at a large software company and a design-oriented job market on Reddit[1]. Seven participants work for an in-house design team either as a creative director ($n = 3$) or a user experience designer ($n = 4$) at a large company; the other three work as full-time freelancers in graphic design. All participants indicated that they receive design feedback regularly as part of their job and reported their age as within 21 to 40 years old. Four study sessions were conducted in person and six were conducted remotely through a video conferencing tool. Each study session lasted approximately 1 hour and the participants were compensated with a $50 gift card.

Procedure

We first conducted a semi-structured interview with each designer where they were asked to describe a recent experience for which they received feedback on a design project from more than one person. We asked the designer to describe how they incorporated that feedback in subsequent revisions (e.g., "How do you decide where to start?" and "How do you resolve contradictions between feedback providers?"). We also asked questions regarding how the designer learned to interpret feedback (e.g., "How has your method of feedback interpretation changed over time?").

Following the interview, each designer participated in a think-aloud feedback interpretation task. We provided the designer with an event marketing flyer (either for a charity concert or a marathon race, randomly assigned) and a document containing a set of feedback written for that flyer, all borrowed from a prior research study [44]. We asked the designers to demonstrate how they would annotate and organize the feedback to devise a revision plan for that flyer. For the charity flyer, the document contained two pieces of expert feedback from Upwork [2], while for the marathon flyer, the document contained seven pieces of novice feedback received from workers in a micro-task market [1]. This setup allowed us to observe if designers adopt different strategies when interpreting feedback written by different numbers of providers or providers with different expertise (though no such differences were observed). Each piece of feedback was about a paragraph long (*mean* = 94 words). For the study sessions conducted in-person, designers were given physical print-outs of flyers and feedback as well as pens and markers so they could annotate feedback or take notes if they wished. In remote study sessions, flyers and feedback documents were shared through Google Docs, and the designer was asked to share their screen as they interacted with the document. All interviews were recorded and transcribed. Observations of the participant behavior and interview transcripts were analyzed through an iterative open coding approach [37].

[1] https://reddit.com/r/DesignJobs/

Formative Study Results

All designers agreed that the ability to interpret multiple pieces of feedback is a critical skill for their work. They also stated that they often receive feedback from multiple people through various communication channels (e.g., email thread, PDFs annotated with comments, Slack, instant message), which makes it difficult to parse various opinions and devise a plan for revision. To manage the process of interpreting multiple pieces of written feedback, we found that experts employed three strategies: identify, categorize, and prioritize.

Identify

Identify refers to the process of distinguishing which pieces of feedback are credible and valuable enough to be addressed.

Upon receiving feedback, designers described judging the value of a piece of feedback based on its perceived insight with respect to the project's overall goals and revision timeline. Designers also considered the status of the person who wrote it (if information about the identity of the feedback provider was available). One designer explained:

> *"The criteria is usually, is this person giving the feedback that he or she is qualified to give? Like I don't really care if a technical advisor doesn't like the color of something."* [P3, F, UX Designer]

Designers also described using self-reflection or asking for further clarification to discern between statements that offer novel insight and those that simply convey the personal preferences of the feedback provider:

> *"I've learned as a designer through practice that oftentimes feedback, especially if it comes from a non-artist, might be based off of a feeling, or they just don't like it. [...] And that's an important distinction, I think, because in the art and design, everybody is going to have opinions."* [P1, M, Creative Director]

Categorize

Categorize refers to the process of organizing feedback into meaningful groups in order to form a high-level view of opinions present in the feedback set.

After identifying the value of each piece of feedback, all designers stated that they maintained a running list of categories to organize feedback content. This list was written down in paper notebooks, Post-it stickers, or stored in software tools (e.g., Evernote, spreadsheets). Some designers recorded feedback verbatim during this organization process, while others

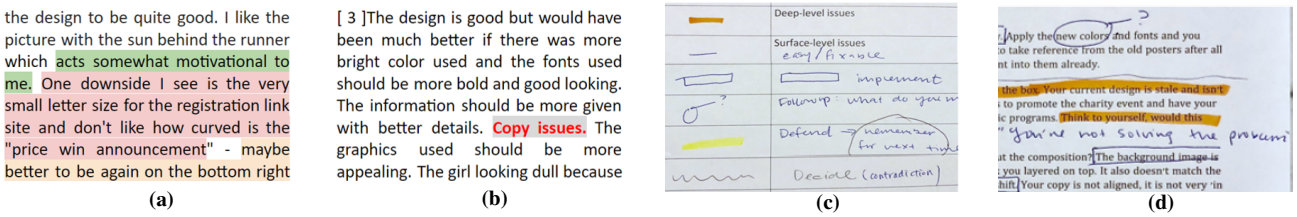


Figure 2: Examples of how experts annotated feedback during the feedback interpretation task in the formative study. (a) The designer used color to highlight the sentiment of the feedback: green (positive), pink (negative), and orange (neutral). (b) The designer inserted a topic label (Copy issues) into the text of the feedback. (c) The designer described her annotations for feedback interpretation; the annotations include topics ("deep-level issues", "easy issues") and intended actions ("implement", "follow-up","defend","decide"). (d) The designer used the schemas in (c) to annotate the feedback.

extracted keywords or a short summary. All of them categorized feedback into groups titled with labels. One designer explained how she structured feedback from multiple people:

> *"The way we have been organizing feedback is on a spreadsheet so that we can put the people that we're communicating with on one column, and then the questions we're asking in rows above so that you can go through each question and say, like, the majority of the people felt this way, and summarize things at the bottom, and so that's been really useful."* [P3, F, UX Designer]

In feedback interpretation tasks, designers demonstrated how they would categorize and summarize the feedback they were given (Figure 2). The types of labels that designers used included the perceived *sentiment* of a piece of feedback, the *topic* or aspect of the design artifact that a piece of feedback related to, and the designer's intended action for a piece of feedback. When categorizing by sentiment, designers stressed the importance of preserving the strength of a design while fixing issues, and included categories in their labeling schemes accordingly.

When categorizing by design topic, some designers used high-level terms (e.g., "easy fix", "deep/conceptual issues"), while others used specific terms describing various aspects of the design (e.g., "copy issues", "hierarchy"). For intended actions, designers marked the statements that they wanted to implement or discuss further with providers. They also highlighted the sentences they disagreed with (e.g., "defend") or needed to think about it (e.g., "decide").

Prioritize

Prioritize refers to the process of deciding what to do first; these decisions may be made based on various factors, including whether feedback aligns with the designer's own opinion, whether there is strong agreement among stakeholders about an idea, the effort required to make a certain change, the authority of the person who wrote a piece of feedback, or the current status of the project. For example, even if a suggested change was an easy fix, designers reported treating that change with low priority if it was perceived as being less relevant for the current status of the project.

Summary of insights

From the formative study, we found that interpreting feedback involved several strategies. These strategies included *identifying* the valuable feedback, designing schema for *categorizing* feedback, and *prioritizing* which feedback to act on. We also found that, while employing these strategies, designers frequently referred to the perceived sentiment and the topic of feedback statements and the status of the provider who wrote the statements. For example, designers would identify critical issues by comparing negative and positive statements regarding a specific issue, categorize feedback by topic, and prioritize possible changes by considering who suggested what.

The strategies designers described were not necessarily performed in a strict order. For example, designers leveraged information about the status of a feedback provider to not just assign value to feedback written by that provider but to also determine its priority for implementation. In sum, we observed that experts created their own structure around free-form feedback in order to effectively perform their work. Despite this, experts reported this process as effortful and time-consuming.

DECIPHER

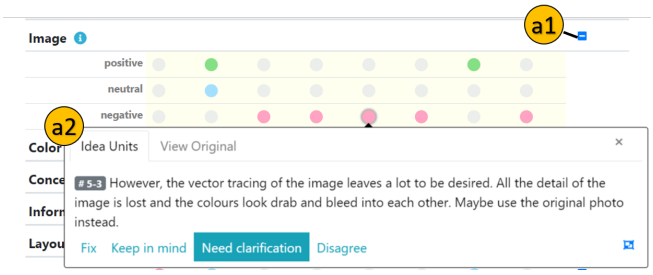
Based on the findings from our formative study, we designed and implemented an interactive tool called Decipher (Figure 1). The tool adds structure to a collection of feedback by visualizing how the feedback maps across providers (the provider of a piece of feedback), topics (categories and themes in the feedback), and sentiment (whether opinions in the feedback are positive or negative). We hypothesize that the representation and interaction in the tool will aid designers in identifying valuable feedback (e.g., by viewing the number of providers that made similar statements) and in prioritizing issues when formulating a revision plan (e.g., by capturing how they want to act on specific statements in the feedback).

In this section, we describe the visualization and interaction mechanisms that comprise Decipher. A design scenario will be used to contextualize the use of the tool.

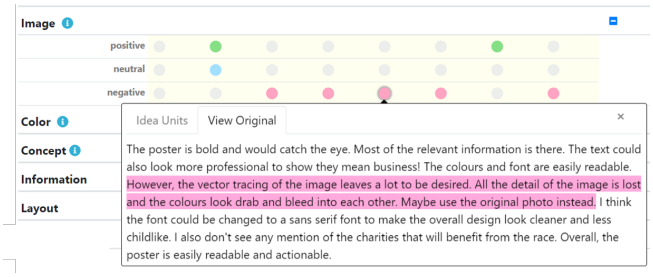
Design Scenario

Imagine a graphic design student, Jun, who has created a prototype of a flyer to advertise a local charity event being organized by a student group at her university. Jun wants to revise the flyer based on feedback from different audiences.

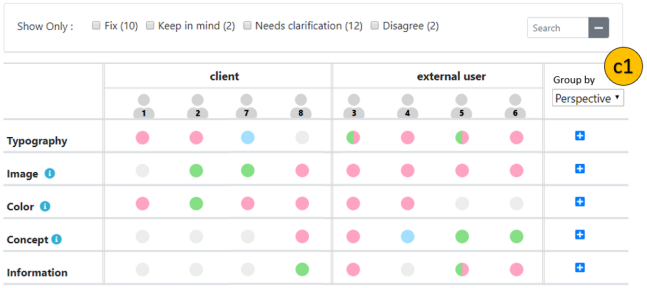
She emails the flyer to her client (the members of the student group) to gain their perspectives as event organizers, and also posts the flyer to an online forum to gain perspectives from people who are potential attendees of the event. Jun receives four pieces of feedback from the student group and another four from the potential attendees, each containing a paragraph or two of text. Jun is initially uncertain in how to revise the design based on all of the feedback she receives because it



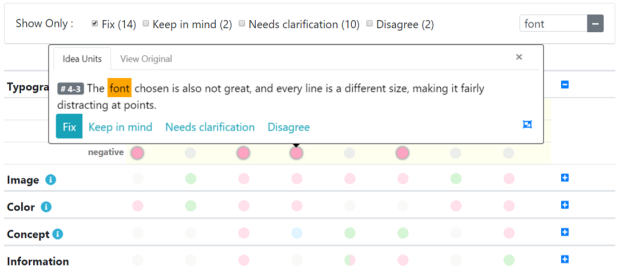
(a) Jun can expand a topic row by clicking the blue plus button (a1) and hovering over each circle to review the feedback details in (a2). Jun can mark the feedback unit using one of the four interpretation labels at the bottom of the window.



(b) Switching to the "View Original" tab will show the unit of feedback (highlighted) in the context of the whole piece of feedback written by that provider .



(c) Jun can group the feedback providers by selecting an attribute in the drop-down list in (c1). In the example, the providers are grouped by their perspective (i.e., client or external user).



(d) The feedback units can be filtered by interpretation labels and keyword. Above, Jun highlights the feedback units that contain the keyword "font" and are marked as "Fix".

Figure 3: User interactions in Decipher.

differs in topic, perspectives, and sentiment, and decides to use Decipher to aid her interpretation of the feedback.

Interpreting the feedback

Decipher visually structures a feedback set by topic and provider and surfaces sentiment patterns with respect to these two dimensions. Using this visualization, a designer can quickly form initial impressions about how the providers perceived the design (e.g., its strengths and weaknesses) without needing to read the content of the feedback directly.

A column in Decipher represents feedback received from one provider, while a row represents all the feedback related to one topic across all providers. A circle represents one or more units within a feedback made by a single provider in reference to a particular topic. Decipher defines a feedback unit as one or more sentences that describe a coherent thought. The color of a circle represents the sentiment of those units: pink indicates negative opinions, green represents positive ones, and blue is for neutral statements or suggestions. If a provider writes both negative (or positive) statements and neutral statements, the circle color will stay pink (or green). However, if a provider writes both positive and negative statements regarding a topic, the corresponding circle will be half-pink and half-green to reflect this tension. If a topic is not present in the provider's feedback, the circle will be light-gray. Topic rows are ordered by the number of non-gray circles they contain, which indicates the popularity of a topic across feedback providers. The metadata used to produce the visualization is collected through a labelling process (see the Constructing Decipher Visualization subsection).

In our design scenario, Jun uses the visualization in Figure 1 to quickly identify the main strengths and weaknesses of the flyer. She expected that the flyer would only require minimal revisions given that she had carefully planned its design. However, Jun is surprised to see many providers expressed negative opinions related to the photograph featured in the flyer (Figure 1, *Image* row). Jun also notices that some providers expressed negative opinions regarding the information present on the flyer while others did not raise concerns about this topic (Figure 1, *Information* row). Finally, Jun notices that the providers in the first and second columns had nearly opposing opinions. To see if the differences of opinion are due to differences in perspective (i.e., design client vs. potential attendee), Jun hovers over the icons of these two providers to view their information. Jun is surprised to see that both providers are clients, and so mentally notes to schedule a meeting with them to confirm her revision plan before actually revising the flyer.

Focusing the interpretation

Decipher provides interactive mechanisms to help a designer to learn more about the high-level feedback patterns they observe.

Drilling into topics. To read the original text of the feedback written about a specific topic, the designer can click on the blue plus icon (Figure 3a-(a1)) to expand the topic row and hover over each circle in that row to display a pop-up window showing the feedback unit corresponding to that circle (Fig-

ure 3a-(a2)). If multiple units correspond to the same circle, Decipher displays all units in a list view. In this window, the designer can click "View Original" to read the units in the context of the original paragraph of feedback (Figure 3b). Here, the units corresponding to the circle of interest are highlighted with their sentiment color.

Organizing the feedback by provider. We saw in our formative study that understanding the background of feedback providers can be important for interpreting the feedback that they write. In Decipher, a designer can organize the columns in the visualization (provider) by attributes such as "perspective" or "expertise" (Figure 3c-(c1)) to place providers with similar backgrounds adjacent to each other. If no attribute is selected, Decipher orders providers by the time each provider submitted feedback, with earlier feedback appearing on the left.

In our design scenario, Jun wants to dig deeper into the negative comments she noticed regarding the photograph used in the flyer. She clicks on the blue plus icon (Figure 3a-(a1)) to expand the *Image* row and hovers over each non-gray circle to read the feedback sentences written by each provider for that topic. Since the visualization shows that there is mixed sentiment among providers with respect to *Image*, Jun organizes the providers by their perspective to explore whether that reveals further patterns in the feedback (Figure 3c-(c1)). Jun found that the event organizers liked her idea of posterizing the flyer photograph (Figure 1), whereas the potential attendees thought this made the flyer (and the event) look unappealing. Jun also notices another pattern she hadn't noticed before: while almost all the potential attendees pointed out that the flyer never actually mentions that the event is for charity purposes, only one of the event organizers noticed that issue.

Capturing interpretations and retrieving them

Decipher enables a designer to capture their emergent interpretation of the feedback. For each feedback unit in the pop-up window, the designer can select any of four interpretation labels (Figure 3a-(a2)):

- *Fix*: used to identify suggestions that the designer agrees should be incorporated in a revision.
- *Keep in mind*: used to identify strengths that should be maintained in a revision or to identify statements that need additional research or reflection.
- *Needs clarification*: used to identify statements that require discussion with the provider who wrote it or with the client in light of the project goals.
- *Disagree*: used to identify suggestions that the designer believes are inconsistent with the project's goals.

Designers can also combine the use of these labels (e.g., to mark a statement for which they currently disagree but want to seek clarification). These initial labels were informed by findings from the formative study on how expert designers annotated their intended actions for the feedback statements, and may be modified in future versions of Decipher as we learn how it aids the interpretation of design feedback.

Filtering feedback. A panel above the visualization allows a designer to filter feedback using keywords or interpretation labels. When one or more labels are selected, or when a keyword is entered in the search bar (Figure 3d), the visualization highlights the circles associated with feedback that matches the filter criteria.

Jun labels the feedback units regarding missing information about the charity event as "Fix" and labels the units that she felt best articulated the issue with the flyer photograph as "Needs clarification" (Figure 3a-(a2)). After capturing these and other interpretations of the feedback, Jun meets with the students organizing the event to discuss the feedback and present her plan for revising the flyer. Jun describes her plan while using the tool to filter for feedback statements labelled as "Fix" and "Needs clarification". She then filters for suggestions labelled as "Disagree" while explaining her rationale for not incorporating these suggestions into the revision. In sum, Jun leverages the Decipher visualization to identify and prioritize issues in the flyer that are in most need of revision, utilizes the interpretation labels to capture her intended actions for specific feedback units, and uses both to discuss her revision plan with her client.

Constructing the Decipher visualization

A labelling process is necessary for generating the Decipher visualization. A designer imports a text file containing the pieces of feedback they receive into Decipher. The tool partitions the feedback into units and presents the designer with an interface for labelling each unit. The current implementation defines a unit as a sentence by default, though the designer can merge multiple related units into larger units during the labelling process. During this process, the designer labels the topic and sentiment for each unit. The interface includes pre-defined topics for the domain of graphic design (e.g., "Typography", "Image", "Concept") and allows the designer to define additional topic labels. The sentiment of each unit can be labelled as positive, negative, and neutral. The designer can also input information about feedback providers such as perspective (e.g., client or external user) and expertise (e.g., novice or expert). This is a purely manual approach for labelling the feedback for use in Decipher. Future work could explore leveraging alternative interaction designs, crowdsourcing workflows, or machine learning techniques to ease the labelling process.

Implementation

Decipher is a Web application built using Javascript, JQuery, HTML/CSS, and the Python Django framework. A PostgreSQL database is used to store feedback content and user-provided labels (e.g., the topic categorizations, attributes of the feedback providers, and the designer's interpretation labels). Nearly all of the tool's features described in the paper have been fully implemented, while a few of the features described (e.g., grouping by providers) are still in-progress.

EVALUATION

We conducted a preliminary evaluation of Decipher to assess how the use of the tool affects how novices interpret design feedback written by multiple providers. We focus on novices in our initial evaluation because we feel this audience currently has the most to gain from tools for feedback interpretation.

Design 1	Design 2
	

Table 1: The flyer designs used in our user study: a Taylor Swift charity concert and a marathon race. Study participants were provided with a flyer design and a set of feedback written for each flyer and asked to interpret the feedback.

In the evaluation, participants completed feedback interpretation tasks with two different representations of feedback—one using the Decipher visualization and the other using raw text in a traditional document editing tool (i.e., Google Docs). The study was designed to answer the following research questions:

- **RQ1**: How does representation of the feedback (Decipher vs. raw text) affect the strategies novices use to perform feedback interpretation?
- **RQ2**: How does feedback representation affect what insights novices do or don't identify in the feedback?
- **RQ3**: How does feedback representation affect the perceived effectiveness of the feedback interpretation process? or interpreting feedback relative to the baseline tool?

We conducted a within-subjects study to control for variation in participants' design expertise, experience with interpreting feedback, and familiarity with the baseline tool. In order to isolate the effect of feedback representation, we provided participants with feedback sets for two graphic designs rather than ask them to interpret feedback individually generated for designs they make.

Participants

Twenty participants (nine female) were recruited via email from two universities in the United States. None of the participants were in our formative study or had advance knowledge of the project. The participants rated their design expertise on a scale from 1 (novice) to 7 (expert). The average self-rated expertise was 2.3 ($SD = 1.4$), indicating that most participants perceived themselves as novices. All participants were between 18-34 years old.

Task Materials

Participants were provided with two event marketing flyers (Table 1) and the associated feedback. The flyers were created by two novices who were not participants in the experiment. The flyers contained text, images, and color and required little domain knowledge to comprehend. Prior to viewing the feedback, participants rated the perceived quality of the flyers on a 7-point Likert scale. Paired t-tests showed no statistical difference between the perceived quality of the flyers (Charity=5.1 vs. Marathon=4.9, *n.s.*).

We recruited eight graphic designers from Upwork to write feedback for each flyer. Each designer had two or more years of professional design experience. We instructed the feedback providers to reference different design topics and discuss both strengths and weaknesses. The goal was to create a set of feedback that differed in focus and opinion. The total length of the feedback collected for each design was similar (around 1300 words). Study participants rated the perceived usefulness of the feedback sets on a 7-point Likert scale. A paired t-test showed no statistical difference between these ratings (Charity=6.2 vs. Marathon=5.8, *n.s.*), suggesting that the feedback was qualitatively similar between conditions.

Tool Conditions

There were two tool conditions: Decipher and the baseline tool. For the *Decipher* condition, we recruited two domain experts to collaboratively label the topic and sentiment of the raw feedback prior to the experiment. These labels are necessary for generating the visualization in the tool. Participants were not required to label the feedback themselves because we wanted to isolate the effect of feedback representation from other aspects of the feedback interpretation process (such as reflection that may occur during the labelling process). Participants viewed a 3-minute demo video of the tool and asked any questions about the tool before the task began.

In the *baseline* condition, participants reviewed feedback in raw text form using Google Docs. Participants were informed that they could leverage any of the tool's features such as highlighting, commenting, and editing the text to aid their interpretation tasks. The sentiment and topic labels collected for the Decipher condition were not provided in this condition.

Procedure

Each participant completed two feedback interpretation tasks, each using a different flyer and study condition. Each task required the participant to review a flyer and the corresponding feedback, then answer questions regarding the feedback.

Review a graphic design. The participant was asked to imagine that they were helping a friend revise a flyer that was intended to be posted in a public place. We oriented the scenario around a design created by a fictional friend in order to reflect the fact that the participant was reviewing a design they did not create.

Review a set of feedback. After reviewing the flyer, we told the participants that their friend collected eight pieces of feedback from graphic designers. The participants were asked to review and summarize the feedback so their friend could later perform the revision without having to revisit the feedback. The participant was then instructed to review the feedback using the tool for their study condition for at least five minutes.

Perform feedback interpretation. Once the participant finished reviewing the feedback, they were asked to write responses to a set of focused questions based on common feedback interpretation goals (e.g., identify key issues/strengths in the design,

locate specific suggestions and find contradictory feedback). To prevent the participant from tailoring their review of the feedback to the questions seen in the first task, we created two variations of the question set and told the participant at the onset of the study that the questions were different for each task. The order of the question sets was randomized for each participant.

Answer a survey. At the end of each task, a survey was administered to understand how the participant perceived the provided feedback (e.g., feedback usefulness) and their experience of interpreting the feedback with the tool.

Finally, after both tasks, the participant completed an exit survey in which they compared Decipher and the baseline tool on a 7-point scale (1=baseline tool preferred, 7=Decipher preferred) across several criteria and wrote free-form responses to explain their ratings; a score above 4 ("Neutral") would indicate a preference for Decipher. The lead researcher then interviewed the participants, asking them to demonstrate how they leveraged the features in each tool to interpret feedback and verbally describe their rationale.

During the study session, the experimenter was available only for resolving technical issues and did not aid the participant in the interpretation task. Each study session was recorded. The combination of the two tool conditions and the two design scenarios were counter-balanced across participants. Each study session lasted roughly 1.5 hours and participants were compensated with $20 via Paypal.

Data Analysis

Open-ended responses were analyzed using an inductive coding approach to develop themes [37]. We also measured the time each participant spent answering the questions about the feedback. In order to compare how well the novices were able to identify critical insights in a set of feedback, we recruited three experts in HCI who had experience teaching design courses to perform the same interpretation tasks using the baseline tool only (to better simulate how they normally interpret and organized feedback). One member of the research team then generated a list of key strengths and weaknesses for each flyer design based on expert responses. We then compared this list with each participant's list of issues identified in feedback interpretation tasks. We opted not to measure feedback interpretation success by capturing and evaluating actual changes to a design because the main contribution of our work is to study behaviors around feedback interpretation. For the exit survey, we performed one-sample t-tests using the neutral rating (4) as the population mean.

RESULTS

All participants ($N = 20$) successfully completed the two feedback interpretation tasks using Decipher and the baseline tool. In the next subsections, we describe the results for each of the research questions. The quotes draw from both the exit survey and the post-interview, we use [**I**: interview; **S**: survey + Participant number] to indicate the source of each quote.

RQ1: Effect of representation on interpretation strategies

Participants reported that the topic categorization in Decipher made the feedback interpretation task less overwhelming compared to reading feedback in the baseline condition ($n = 13$). They attributed this to the ability of processing feedback in shorter and more focused segments, one aspect at a time:

> *"If you look at this [Decipher] and then take a look at this [the baseline tool], they're technically the same thing; and the good thing about the tool [Decipher] is that for each feedback, you are just reading one or two sentences at a time. You won't feel overwhelmed when looking at two sentences compared to a five, six sentence long paragraph."* [**I**1015, Charity (Decipher) + Marathon (baseline)]

The grid-based representation of feedback also provided a scaffold for our novice participants to process feedback by topic rather than by its written order, which resembled the behavior we observed from experts in the formative study. We noticed that all the participants followed the topics sequentially–they all started from the top row and browsed most of the feedback units in each row by hovering over circles.

In contrast, with the baseline tool, participants reported that reading and processing feedback required extra cognitive effort because feedback, even from a single provider, typically mentions multiple aspects of the design. To facilitate the interpretation process, eighteen participants categorized the feedback by highlighting or commenting statements that were important to them; however, many of them mentioned that their attention gradually dropped over time, with five participants stating that they skipped text in the document. One participant described how the lack of engagement, combined with the sense of being overwhelmed, negatively affected their ability to process and highlight feedback using the baseline tool.

> *"You can see how I was probably doing everything on top, and then as I go down [...] I tend to lose focus since I'm reading a lot of paragraphs and stuff. You would still do them, but your organization might not be accurate."* [**I**1014, Marathon (Decipher) + Charity (baseline)]

In addition to the topic categorization, most participants ($n = 14$) reported that they used the color-coded sentiment to help them determine the importance of each topic:

> *"The tool (Decipher) helps me recognize the importance of each issue by seeing how many people pointed to the same issue. When 6 out of 8 people comment on the font issue, then it is pretty clear that the font has to be modified."* [**S**1040, Marathon (Decipher) + Charity (baseline)]

Using the baseline tool, although participants reported that they mentally used the same criteria for determining feedback importance, many of them expressed the difficulty of developing suitable topic labels for categorizing issues:

> *"After reviewing first two or three pieces of feedback, I started to see that people mentioned similar stuff [...] I need to come up with tags to categorize the text, but it's hard for me to generate those categories. That's why I*

did not highlight all the things even I think are important." [**I**1017, Charity (baseline) + Marathon (Decipher)]

The sentiment visualization also provided context that better prepared our novice participants for processing contradictory feedback ($n = 6$):

"The feedback was nicely categorized by sentiment. It made it easier to read as my mindset is already expecting particular criticism or praise on the topic, it's like I can be mentally prepared for the upcoming information." [**S**1026, Marathon (Decipher) + Charity (baseline);]

With the baseline tool, participants reported that it was harder to process such contradictions as the related pieces might be scattered around in the document.

Finally, Decipher acted as an interactive menu for exploring the feedback space. Participants mostly used the row labels (i.e., topics) to target feedback they were interested in or wanted to refer back to. With the baseline tool, they typically used keyword search (Ctrl+F) or scrolled the page.

RQ2: Effect of representation on identified insights

Decipher helped novices identify critical issues more effectively than the baseline tool when these issues were popular across the feedback providers. In our study, among the four critical issues (two per flyer) that were identified by the experts, three of them were mentioned by more than half of the feedback providers. We found that all but two Decipher participants were able to discover those three issues (e.g., "irrelevant borders" in the charity flyer and "the artificial effect made to the background photo" in the marathon flyer). In contrast, participants in the baseline condition missed those issues more often; for example, although five providers suggested "remove the irrelevant border" in the charity concert flyer, less than half of the baseline participants considered it as a critical issue.

Participants in the Decipher condition were also more sensitive to conflicting opinions than in the baseline condition, especially when the discrepancy between the amount of positive and negative feedback was large. For example, three providers liked the choice of the color palette used in the marathon flyer whereas one provider thought it was drab. While all the Decipher participants and the experts spotted the contradiction, less than half of baseline participants captured that.

However, Decipher participants were likely to miss critical issues that were not mentioned frequently by providers. For example, in the marathon flyer scenario, the experts reported that "missing charity vibe" was a critical issue present in feedback even though it was mentioned only by three out of eight providers. In this case, roughly half of the participants missed the issue regardless of the tool used.

RQ3: Effect of representation on user perceptions

According to survey responses, participants preferred Decipher over the baseline tool ($M = 5.8$, $SD = 1.3$; $t(19) = 4.56$, $p < 0.001$). In particular, participants reported that Decipher made it easier for them to identify critical issues ($M = 5.7$, $SD = 1.3$; $t(19) = 8.0$, $p < 0.001$) and locate specific feedback ($M = 6.3$, $SD = 1.3$; $t(19) = 6.05$, $p < 0.001$). In addition, most participants reported that Decipher helped them comprehend feedback as well or better than the baseline tool ($M = 4.7$, $SD = 1.5$; $t(19) = 1.99$, $p = 0.03$) while making them feel less overwhelmed due to the amount of feedback ($M = 3.2$, $SD = 1.5$; $t(19) = -1.87$, $p = 0.038$). All participants anticipated that Decipher would make it easier to discuss feedback with others ($M = 6.2$, $SD = 0.9$; $t(19) = 9.18$, $p < 0.001$).

Finally, Decipher did not seem to have a negative effect on participants' engagement with feedback relative to the baseline tool. A paired t-test suggested that the participants spent a similar amount of time processing feedback in both conditions (Decipher= 21.7 minutes v.s baseline= 19.5 minutes, $t(19) = -1.6, n.s.$). On average, participants used 8.8 minutes ($SD = 4.5$) to review feedback in Decipher and 8.4 minutes ($SD = 2.8$) in the baseline tool, which was not significantly different between the conditions ($t(19) = -0.4, n.s.$). Similarly, the participants spent 12.9 minutes ($SD = 6.6$) in Decipher and 11.2 minutes ($SD = 4.4$) in the baseline tool to answer questions about the feedback in each design scenario ($t(19) = -1.3, n.s.$).

While these findings should be considered preliminary, we see encouraging evidence that 1) Decipher helped novices find the same issues experts identified in a collection of feedback, 2) Decipher participants paid attention to both strengths and weaknesses described in feedback, and 3) Decipher reduced the sense of being overwhelmed while processing a collection of feedback. On the other hand, we also see evidence that Decipher caused participants to miss issues when they were not commonly mentioned and explicitly emphasized by the visualization (for example, the "vibe" of the Charity poster).

DISCUSSION AND FUTURE WORK

Decipher leverages interactive visualization to aid feedback interpretation. Through a controlled study, we found that Decipher helped novice designers feel less overwhelmed, find patterns, and identify contradictions while processing a set of detailed text feedback from multiple providers. In this section, we discuss design implications for systems that support feedback interpretation and propose areas for future research.

In the comparative study, the feedback in the Decipher condition was pre-annotated to allow participants to focus on reviewing feedback in the visualization. In practice, designers would need to perform this step themselves. Future work is needed to test how effectively designers are able to perform this annotation task in the context of a complete feedback loop. However, we do not believe that the need for categorizing feedback necessarily diminishes the value of a tool like Decipher. We view this trade-off as analogous to how note taking, while requiring a student to take the notes, is known to have significant learning benefits [27, 26].

Future work can also explore how to reduce annotation effort through crowdsourcing and natural language processing techniques. For example, research has shown that online crowds can produce annotated data sets that are as accurate as those produced by experts given the proper schemas and guidance [41, 33]. Crowds can also coordinate to generate schemas to reflect emergent topics in design feedback or other domains [8, 3, 18]. Topic modelling techniques such as LDA [5] can gener-

ate topics based on the data, but may require a hybrid approach where the crowd (or user) names the algorithm-generated categories or an interactive approach where the user specifies the type of feedback they are interested in. This flexibility would help generalize the mechanisms implemented in Decipher to domains other than design, which each have their own set of terminology and language for feedback.

Our evaluation studied the use of Decipher in a controlled context where users performed a single round of feedback interpretation as part of standalone design tasks. However, in reality, designers typically receive multiple rounds of feedback on a project [36] and are often interested in gauging how their skills improve over time [34]. In future work, an interface for viewing and comparing multiple visualizations of feedback received on a project may make it possible to see how a project's strengths and weaknesses have changed with iteration (at least, according to the feedback). This may also allow designer to see how their skills have evolved across several projects, enabling a deliberate reflective practice that is an essential aspect of gaining mastery in a domain [12].

Participants reported in the exit survey that visualizing feedback by topic reduced the sense of feeling overwhelmed because they could process feedback in focused segments. Along these lines, future work is needed to tease apart how specific features of Decipher support (or hinder) feedback interpretation. For example, one design choice we found contentious among study participants was the presentation of sentiment. Some participants liked sentiment as a primary element of the Decipher visualization because it allowed them to quickly realize the weaknesses and strengths of the design, but other participants found it discouraging to see the prominence of negative sentiment in the tool. In future work, rather than presenting feedback as generally positive or negative, orienting feedback around concrete concepts such as "changes to consider" or "successes to maintain" may better help novices mimic the way experts think about feedback.

Another open question is how the way feedback is presented could facilitate feedback exchange. It may be useful for feedback providers to see the same visualization of the feedback that the designers would see. Feedback providers could use Decipher as a map of feedback that has already been given or to compare the attitude of their feedback (as interpreted by the tool) with those of others and adjust the scope and content of their feedback accordingly. Prior work shows that individuals are more likely to contribute feedback online if they believe they can make a meaningful contribution [28]. The visualization could also act as a channel for designers to communicate their interpretation and intended actions of the feedback back to providers (e.g., through interpretation labels).

The current implementation of Decipher allows designers to label intended actions for feedback statements and filter the feedback based on these labels. Additional features could further aid generating a revision plan. For example, a designer could enter time estimates for the *Fix* labels and generate possible revision plans given different timelines. Future work is needed to test how a tool like Decipher could best assist designers when thinking about the next steps for their work.

While the design of Decipher was grounded in formative observations of expert practices, our comparative study primarily focused on how novices use Decipher to navigate feedback. Given that experts from our formative study commented that they find current practices of feedback interpretation tedious, we expect Decipher will make interpretation easier for experts as well. Comparing how novices and experts use a tool like Decipher also presents an exciting opportunity for future work.

In addition to addressing the issues already discussed, we see several directions for future work. One direction is to compare and contrast how users interpret feedback in different domains (e.g., for research proposals, course instruction, or job performance) to expand the strategies reported in this paper. A second direction is to expand the tool for use with large-scale feedback sets (e.g., the large volume of feedback that might be received from visitors to a Web page through a link placed on the page) and implement features to help with preparing an effective revision plan. Finally, future work is needed to test how designers might use Decipher at different project stages and for different kinds of creative work.

LIMITATIONS

One limitation of our work was that participants were provided with existing designs and feedback in the interpretation tasks. We imagine that the participants might react differently if they were interpreting feedback on their own designs. For instance, a designer might be more surprised by contradictory feedback or feel more discouraged by criticisms. Future work is needed to test how emotional responses might affect the interpretation of feedback and to learn how a tool like Decipher may influence the way people interpret feedback and plan the next steps for their own work. Another limitation was that the participants did not actually revise the design. Future work could monitor if and how the interpretation of feedback changes over the entire revision process; for example, a designer may deprioritize an intended change after realizing that incorporating the change is more difficult than originally thought.

CONCLUSION

Creative feedback is difficult to write and even more difficult to understand, especially when it involves resolving contradictions across multiple perspectives, judging the credibility of suggestions, and prioritizing the multitude of issues that are raised. To better understand how to guide feedback interpretation, an under-supported component of the creative process, we conducted a formative study with ten design professionals and identified three strategies they use to interpreting feedback they receive from multiple sources: Identify, Categorize, and Prioritize. Based on these strategies, we then designed Decipher, a visualization tool to help novice designers interpret sets of feedback from multiple sources. Our work is a first step towards helping creators benefit more effectively from the suggestions, encouragement, and knowledge they receive from peers, mentors, and clients.

ACKNOWLEDGMENTS

This work was supported in part by NSF awards 1462693 and 1530818 and a graduate dissertation fellowship awarded to the first author.

REFERENCES

[1] 2019. Amazon. (2019). https://amazon.com.

[2] 2019. Upwork. (2019). https://www.upwork.com/.

[3] Paul André, Aniket Kittur, and Steven P. Dow. 2014. Crowd Synthesis: Extracting Categories and Clusters from Complex Data. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 989–998. DOI: http://dx.doi.org/10.1145/2531602.2531653

[4] Frederik Anseel, Filip Lievens, and Eveline Schollaert. 2009. Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes* 110, 1 (2009), 23–35.

[5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[6] Kirsten R Butcher and Tamara Sumner. 2011. Self-directed learning and the sensemaking paradox. *Human–Computer Interaction* 26, 1-2 (2011), 123–159.

[7] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman (Eds.). 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[8] Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. 2013. Cascade: Crowdsourcing Taxonomy Creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1999–2008. DOI: http://dx.doi.org/10.1145/2470654.2466265

[9] Amy Cook, Jessica Hammer, Salma Elsayed-Ali, and Steven Dow. 2019. How Guiding Questions Facilitate Feedback Exchange in Project-Based Learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 138, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300368

[10] Patrick A. Crain and Brian P. Bailey. 2017. Share Once or Share Often?: Exploring How Designers Approach Iteration in a Large Online Community. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (C&C '17)*. ACM, New York, NY, USA, 80–92. DOI:http://dx.doi.org/10.1145/3059454.3059476

[11] Stanford d.school. 2019. Design method: I like, I wish, what if. (2019). https://dschool-old.stanford.edu/wp-content/themes/dschool/method-cards/i-like-i-wish-what-if.pdf.

[12] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological review* 100, 3 (1993), 363–406.

[13] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1175–1184.

[14] Eureka Foong, Steven P. Dow, Brian P. Bailey, and Elizabeth M. Gerber. 2017a. 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)*. Association for Computing Machinery, New York, NY, USA, 4454–4467. DOI: http://dx.doi.org/10.1145/3025453.3025791

[15] Eureka Foong, Darren Gergle, and Elizabeth M. Gerber. 2017b. Novice and Expert Sensemaking of Crowdsourced Design Feedback. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 45 (Dec. 2017), 18 pages. DOI: http://dx.doi.org/10.1145/3134680

[16] C. Ailie Fraser, Tricia J. Ngoon, Ariel S. Weingarten, Mira Dontcheva, and Scott Klemmer. 2017. CritiqueKit: A Mixed-Initiative, Real-Time Interface For Improving Feedback. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 7–9. DOI:http://dx.doi.org/10.1145/3131785.3131791

[17] 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. DOI: http://dx.doi.org/10.1145/2757226.2757249

[18] Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. 2016. The Knowledge Accelerator: Big Picture Thinking in Small Pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2258–2270. DOI: http://dx.doi.org/10.1145/2858036.2858364

[19] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research* 77, 1 (2007), 81–112.

[20] Amy J. Henley and Florence D. DiGennaro Reed. 2015. Should you order the feedback sandwich? Efficacy of feedback sequence and timing. *Journal of Organizational Behavior Management* 35, 3-4 (2015), 321–335. DOI: http://dx.doi.org/10.1080/01608061.2015.1093057

[21] Pamela J Hinds. 1999. The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied* 5, 2 (1999), 205–221.

[22] Jeff Huang, Oren Etzioni, Luke Zettlemoyer, Kevin Clark, and Christian Lee. 2012. RevMiner: An Extractive Interface for Navigating Reviews on a Smartphone. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. Association for Computing Machinery, New York, NY, USA, 3–12. DOI: http://dx.doi.org/10.1145/2380116.2380120

[23] Maria Jackson and Leah Marks. 2016. Improving the effectiveness of feedback by use of assessed reflections and withholding of grades. *Assessment & Evaluation in Higher Education* 41, 4 (2016), 532–547.

[24] Anders Jonsson. 2013. Facilitating productive use of feedback in higher education. *Active learning in higher education* 14, 1 (2013), 63–76.

[25] 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, Article 606, 13 pages. DOI: http://dx.doi.org/10.1145/3173574.3174180

[26] Kenneth A Kiewra. 1985. Students' note-taking behaviors and the efficacy of providing the instructor's notes for review. *Contemporary educational psychology* 10, 4 (1985), 378–386.

[27] Kenneth A Kiewra, Nelson F DuBois, David Christian, Anne McShane, Michelle Meyerhoffer, and David Roskelley. 1991. Note-taking functions and techniques. *Journal of educational psychology* 83, 2 (1991), 240–245.

[28] Robert E Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design*. Mit Press.

[29] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer. 2013. Peer and Self Assessment in Massive Online Classes. *ACM Trans. Comput.-Hum. Interact.* 20, 6, Article 33 (Dec. 2013), 31 pages. DOI: http://dx.doi.org/10.1145/2505057

[30] Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*. ACM, New York, NY, USA, 342–351. DOI: http://dx.doi.org/10.1145/1060745.1060797

[31] Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A. Myers. 2019. Unakite: Scaffolding Developers' Decision-Making Using the Web. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 67–80. DOI: http://dx.doi.org/10.1145/3332165.3347908

[32] Weichen Liu, Sijia Xiao, Jacob T. Browne, Ming Yang, and Steven P. Dow. 2018. ConsensUs: Supporting Multi-Criteria Group Decisions by Visualizing Points of Disagreement. *Trans. Soc. Comput.* 1, 1, Article Article 4 (Jan. 2018), 26 pages. DOI: http://dx.doi.org/10.1145/3159649

[33] 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. DOI:http://dx.doi.org/10.1145/2675133.2675283

[34] Kerry Platman. 2004. 'Portfolio careers' and the search for flexibility in later life. *Work, employment and society* 18, 3 (2004), 573–599.

[35] Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. ACM, 269–276.

[36] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.

[37] David R Thomas. 2003. A general inductive approach for qualitative data analysis. (2003).

[38] Karl E Weick, Kathleen M Sutcliffe, and David Obstfeld. 2005. Organizing and the process of sensemaking. *Organization science* 16, 4 (2005), 409–421.

[39] Naomi E Winstone, Robert A Nash, James Rowntree, and Michael Parker. 2017. 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience. *Studies in Higher Education* 42, 11 (2017), 2026–2041.

[40] Y. Wayne Wu and Brian P. Bailey. 2018. Soften the Pain, Increase the Gain: Enhancing UsersâĂŹ Resilience to Negative Valence Feedback. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article Article 186 (Nov. 2018), 20 pages. DOI:http://dx.doi.org/10.1145/3274455

[41] 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. DOI: http://dx.doi.org/10.1145/2531602.2531604

[42] Koji Yatani, Michael Novati, Andrew Trusty, and Khai N. Truong. 2011. Review Spotlight: A User Interface for Summarizing User-Generated Reviews Using Adjective-Noun Word Pairs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1541–1550. DOI: http://dx.doi.org/10.1145/1978942.1979167

[43] Yu-Chun (Grace) Yen, Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. 2016. Social Network, Web Forum, or Task Market?: Comparing Different Crowd Genres for Design Feedback Exchange. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 773–784. DOI: http://dx.doi.org/10.1145/2901790.2901820

[44] Yu-Chun Grace Yen, Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. 2017. Listen to Others, Listen to Yourself: Combining Feedback Review and Reflection to Improve Iterative Design. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (C&C '17)*. ACM, New York, NY, USA, 158–170. DOI: http://dx.doi.org/10.1145/3059454.3059468

[45] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Bjorn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1005–1017. DOI: http://dx.doi.org/10.1145/2818048.2819953