

Dear Reviewers & Committee Members,

Thank you for the helpful and constructive feedback. We have taken all of this feedback into account in revising our paper. In brief, we added 2 new subsections, 2 new figures, and made considerable modifications to every section of the paper to provide more insights into our design process and findings. We attach a detailed summary of changes that we have made below. We organize the summary into sections corresponding to required changes, recommended changes, and other changes. We have also labeled each item with the reviewers who raised these concerns and quoted the reviewers (in italics) when appropriate. We believe that the listed changes have substantially improved the clarity and readability of our revised paper.

Sincerely,
The Authors

Required Changes

(1) Paper confusing structure due to referencing future sections (R1, R2, R4)

We agree that the paper text often references future sections; this stems in part from the non-traditional format of our paper that employs a multi-domain user-centered design approach and evaluation. In our revision, following up on the reviewers' feedback, we have made our paper easier to read by minimizing references to future sections. This has led to the following changes:

- In Sections 3, we eliminated references to Section 4, 5 and footnote references to features that have not yet been discussed
- We moved the Table of key components in VQSs (now Table 3) from the related work section (S2) to the findings section (S7), after the sensemaking processes and components have been introduced in Section 6.
- We have also deleted three figures (combining Figure 3 taxonomy with Table 2, removing Figure 1 and 4) to minimize the numerous unnecessary cross-references between the text and figures at the beginning of Section 6 and throughout Section 7.
- We have also removed the references to the Appendix to make the paper more self-contained: in Section 3 (removing reference to Appendix A), Section 5 (removing references to Figure 8 in Appendix), and Section 6 (removing reference to Appendix B).

(2) More details regarding participatory design process, while removing lengthy justifications for various aspects of the study design (R1, R2, R3, R4)

To accommodate space for more details regarding the design process, we have removed the bulk of the related work introducing Pirolli and Card's sensemaking model in Section 2 (deferring only the necessary description to Section 6). We now describe our approach as user-centered design, in accordance with Sharp, Preece and Rogers [47], where "*all design decisions are taken*

within the context of the users, their work, and their environment". Indeed as R2 suggests, our approach is better described as a user-centered approach that includes contextual inquiry and participatory design. This shift in terminology has resulted in the removal and modification of approximately four-paragraphs worth of content towards the end of Section 2 and throughout Section 3, resolving the concern for "*lengthy justifications of the utility of contextual inquiry and participatory design*" (R2). We also removed the introduction to grounded theory coding methods in Section 5.1.

Using this space, we added four paragraphs to shed more light on "*what actually happened*" (R2) during the participatory design process. We described advantages and shortcomings of the collaborative feature discovery process, such as how parallel engagement with multiple domains leads to cross-pollination of feature designs and why certain features were not carried from the design to implementation stage. We also added more details:

- To address the "general confusion in terms of methodological terminology" (R1), we clarified our approach into three separate phases, indicated by subsection titles in Section 3 and the further illustrated in the new Figure 1 (the lifecycle model of different design activities).
- We described the focus of the contextual inquiry and details of interactions with participants and what they did during the collaborative process (S3.1)
- We provided details on how we iterated on the design of the prototype with participants (S3.21)
- We provided "*greater exposition of the contextual inquiry*" (R2) by adding more information regarding participant's existing workflows, including the addition of Figure 3 with screenshots captured during contextual inquiry (S4.1-4.3)
- Added a new subsection summarizing the themes that emerged from the need-finding phase (S4.4)
- Added a new section in the appendix containing further details on the evaluation study protocol, including the pre- and post-study survey questions and the semi-structured interview questions during the main experiment (Appendix C)
- Provided additional details on A3's use case from the evaluation study (S7.1)
- Added a new section in the appendix containing further details on the evaluation study protocol, including the pre- and post-study survey questions and the semi-structured interview questions during the main experiment (Appendix C)

(3) Quantitative analysis inappropriate mechanism for reporting study results. (R2, R4)

We agree with the reviewers that our study is "*more like a case study than a traditional experiment*", so rigorous reporting of quantitative metrics is not as appropriate. The reporting of quantitative findings in our study was included as a triangulation point to see if the findings paralleled our qualitative findings and as a prompt for future research questions. They are not our central contribution. As such, we replaced some of the quantitative statistics, such as averages reported for the number of years of experience, number of meetings per month, participant gender breakdown (S3.1), and length of evaluation study (S3.2) with more

qualitative descriptions. We better articulated the qualitative nature of our study in S7.1. Further, as suggested by R2, we removed the original Figure 4 (bar chart of different ways to generate pattern queries) and its mentions throughout S7. We added a citation for the Markov model to highlight how such analysis affords both qualitative and quantitative insights (S7.3). Whenever we report statistics (such as frequencies of various operations, which is essential for some of our findings), we have added text acknowledging the limitations and issues with interpreting the numbers as generalizable metrics. We have also noted this in the Limitations section in S7.4.

Recommended Changes

(4) Disconnect between referenced methodology and actual approach (R2, R4)

We mitigated this issue by shortening the related work regarding participatory design in S2 and motivating these methodological choices in S3 to establish the connection with our work. Given that many reviewers have described our work as user-centered design, as well as R4's explicit concern regarding our similarity with user-centered design, as described above in our response to (2), we described our process in detail within a user-centered design framework.

(5) Discuss issues related to the “*parameter selection problem for the VQS analytics engine*” (R3)

We added a paragraph in S7.1 on how users change analytical parameters for match refinement, and while we acknowledge that this is not the focus of our study, we offer some preliminary recommendations for future investigation.

We made many other minor modifications to the paper. Specifically, we:

- Modified Table 3 (now Table 2) with additional column based on sensemaking process (previously only color coded) to address R1's concern that it is difficult to know without reading the caption what the colors signify (R1)
- Made a minor addition to video to emphasize the importance of such design studies and how it is often overlooked in the VIS community (1:15-1:30) (R2)
- Enlarged inline figures in S7, Figure 6, Table 3 for legibility (R2)
- Fixed subsection heading to standard spacing (R2)
- Fixed minor typographic/grammatical errors: (R2)
 - p.1, paragraph 2: “...how prior work HAS focused...”
 - p.1, paragraph 4: “These missing capabilities partially EXPLAIN...”
- Clarified how the Markov Model captured the % of times they transitioned from process X to Y, rather than the absolute proportion of total time spent in S7.3(R2)
- Added a footnote justifying why S7.3 findings were broken down by domain rather than by individual (R4)
- Clarified that the evaluation study is performed in-lab, rather than in the field, in S3.2 (R4)
- Removed the confusing footnote that clarifies the synonymous use of time series and line charts in S2 (R4)