

# Towards Identification and Mitigation of Task-Based Challenges in Comparative Visualization Studies

BELIV Workshop Submission Paper ID: 1854

Aditeya Pandey, Uzma Haque Syeda, and Michelle A. Borkin

We thank all the reviewers for your time and invaluable feedback. In addition to minor wording edits/corrections, we incorporated the reviewer's comments into this new submitted version. Below you will find the original comments from each reviewer and a brief explanation of how we addressed each comment.

In addition to the report we have also uploaded an annotated version of a revised paper that shows significant changes. A sample text from the Annotated PDF is shown in figure "Annotated Document". The legend to read the Annotated Document is as follows: the **red text** shows the removed content, and the **blue text** indicates the added content to the final submission.

We worked hard to address all the reviewers' helpful comments and be as transparent as possible with our responses. However, if we have missed out on something important that warrants further revisions, please feel free to reach out to us.

Sincerely,

Aditeya Pandey, Uzma Haque Syeda, and Michelle A. Borkin

---

## Primary Reviewer

However, some negatives were identified that must be addressed in the paper as a condition for acceptance:

(-) The precise meaning of "task abstraction" is not clear: are there existing taxonomies that do task abstraction, or is something different from what those taxonomies do intended in this paper? See R2's comment on C2. This should be clarified.

**Response:** We thank the reviewer for highlighting the ambiguity in our task abstraction explanation. In this paper, we refer to the task abstraction concept discussed by Tamara

Munzner in her book Visualization Analysis and Design (VAD). Munzner defines task abstraction as: “Transforming task descriptions from domain-specific language into abstract form allows you to reason about similarities and differences between them.” To clarify the confusion, we have elaborated on the definition of task abstraction and cited the VAD book for further reference in Section 2.2 (Related Works: Task Abstraction and Task Space). We also realized that our example of task abstraction needs to be more general. Therefore we replace our current task abstraction example that uses a tree visualization task with the example used by Tamara Munzner in the VAD book. We believe the updated task abstraction will be more relatable to the readers. Additionally, in section 5 challenge C2 we added a reference to the related work section and a brief definition of task abstraction to reiterate our understanding of task abstraction.

(-) Many of the papers in the survey are older, and some of the same challenges may not apply to newer papers (R2 on C1). It would be helpful to comment on this, even if only as a caveat if limited evidence is available from the existing dataset.

**Response:** To address this feedback, we have included a caveat in our discussion (Section 5) about our survey’s scope and focus. More specifically, we have added a limitation that our evidence is preliminary, and further evidence may be required due to the extent of our survey. We have also discussed how new evidence can be collected to validate the challenges.

(+/-) The paper is better categorized as a Survey paper than a Position paper (R1). I agree, and have updated the categorization in PCS. The authors should make sure their paper reflects this as well (I believe the only difference is that the subtitle "Position paper" will not appear on the paper in the proceedings).

**Response:** We thank the reviewer for changing the category of our paper to Survey. Going into the submission, we also had doubts about the type of paper. Most of our writing was styled in a survey paper style. The challenges that we present were also identified through a review of the literature. Therefore, we are convinced by arguments of R1 and believe the paper is poised best to be a survey paper. To ensure that our paper reflects the qualities of a survey paper type, and is not mistaken as a position paper, we have:

1. Replaced the phrase “position” with “survey” in the Introduction. We have also checked for the words “position” in the rest of the article.
2. Added the keyword “survey” in the abstract. Specifically, we have added: “The challenges discussed in this paper are further backed by evidence identified in a detailed survey of comparative tree visualization studies.”
3. We have hosted the supplemental materials of the survey on osf (<https://osf.io/g3btk/>) to ensure that other researchers can analyze the survey results in detail.

We also reviewed other sections of the paper and found that overall structure and writing of the already emphasizes that our work is a survey of task-based challenges.

Optionally, this point should be addressed:

(-) Section 3 breaks the flow of the paper without being crucial to it, and could possibly be removed (R1). I (meta) am inclined to agree: it is primarily only Stages 3 and 6 (study design and paper writing) from section 3 that are referenced anywhere else in the paper. One approach might be to discuss the relevant information from these stages in context, perhaps keeping the figure but leaving out the long-form prose description of each stage (which should anyway be familiar to any researcher reading the paper).

**Response:** We have incorporated the helpful suggestions recommended by the reviewers. We have removed Section 3 (A Preliminary Life cycle of an Evaluation Project in Visualization) but kept Figure 2 in the paper. In Section 5, where we discuss challenges, we also replaced the reference of Section 3 with an in-context discussion of each problem. We refer to Figure 2 for additional context about the evaluation project's life cycle.

The remaining points raised by both reviewers I take more as possible points of discussion or provocations that could be responded to within the paper if the authors desire (e.g., R1's question of whether or not "tasks" are even the right "atomic thing to study in our evaluations"). I do not consider these crucial points to address, but discussing them would likely enhance the paper, so I encourage the authors to consider the reviewers' remaining points carefully in revision.

**Response:** We appreciate the help provided by the primary reviewer in clearly identifying the critical discussion points. For the comments and issues raised by other reviewers, we provide direct response to their feedback in the next two sections of this document and indicate relevant changes made in response in the paper.

Finally, R1 provides a compelling argument that the paper is more of a survey paper than a position paper, and provides concrete suggestions for how the paper might be pushed from the "survey" to "position paper" category, in their view.

**Response:** Following the consensus with R1, we chose to solidify our paper as a survey paper which is a better fit than "position". As a result, we will focus on making this paper

a solid survey paper and not incorporate suggestions to strengthen it as a position paper.

---

## Reviewer 1

1) The authors claim this submission type is a position paper. I disagree. I think it is a survey paper, since it is centered around the survey, and the positions taken up by the authors are either too general to really brook disagreement, or are not really defended against counter-positions. The survey directly motivates the intended contribution of the work, to me. This is my primary reason for my less than perfect score: when judged by the criteria of a position paper I feel this doesn't quite meet the mark. As a survey it's above and beyond.

**Response:** We have responded to this point in the primary reviewer's comments.

2) I feel like there is some deadweight in the paper. Section 3 (and Figure 2) describe an evaluation in a sort of idealized form that breaks up the flow of the paper and does not seem that crucial to the paper itself. Figure 3 also seemed like something that could just be a sentence or two of text (if it is retained as a figure, the authors should explicitly indicate the TDAI/TDA/TD acronyms in the figure and/or caption, rather than relying on the bullet list in 5.3)

**Response:** We have removed Section 3 from the paper as discussed more specifically in response to the primary reviewer's comment. We argue that Figure 3 is crucial to communicate the concept of evaluation tasks, abstraction of tasks and task instructions to our readers, therefore we plan to retain Figure 3. But, as suggested by the reviewer, we have added labels (TDAI/TD) to sub-figures to improve their readability.

3) I was generally on board with the nature and scope of the problem as laid out by the authors, but I was less convinced by the proposed solutions. Some potential devil's advocate ideas I had while reading the paper:

3A) Maybe "tasks" are just the wrong category of atomic thing to study in our evaluations. After all, most of the "tasks" laid out in the paper are things that could be solved through a table lookup and have one unambiguous answer. Isn't the promise and power of visualization supposed to be for higher level analytical and syncretic goals? So rather than reforming tasks to fit into the abstractions laid out by the authors and in other work, we should just give up on the whole model and choose something else (for instance, the "personas" concept used in industrial user research)

**Response:** We thank the reviewer for sharing their view on the transition of visualization evaluation focus from low-level analytical tasks to high-level visualization goals. We genuinely believe this is an essential question for the community and will be very

beneficial to discuss in a workshop like BELIV. However, in our current paper we did not find a place to discuss this idea without introducing unnecessary doubts among our readers. Although we do not directly address this comment in the paper, we believe similar questions will be raised in the QA session of the paper and will lead to some interesting debate within the community.

Furthermore, we do not entirely agree with the view of the reviewer. We believe visualization evaluation studies should focus on **both** low-level and high-level visualization goals. From the generalizability of visualization knowledge purpose, we are convinced that low-level analytical tasks are irreplaceable and will continue to influence visualization knowledge.

Although outside the scope of our paper, we look forward to future publications and discussions at BELIV on the role and importance of high versus low level tasks.

3B) <https://xkcd.com/927/> : the authors attempt to make everybody play nice might create more problems down the line. The large numbers of task abstractions and description standards laid out by the authors point to me that there has been tremendous duplication of effort in the field, and so

3C) Do I necessarily /care/ that all of these task comparisons are apples to oranges? <https://treevis.net/> lists 318 or so different tree visualization types, and I'm not certain it's worth our time as a field to quantitatively evaluate each one against each other in a comparable way, especially since they are often deployed to domain experts who would have idiosyncratic information-seeking habits even /if/ we were all on the same page task-wise.

**Response:** We believe the reviewer is referring to Challenge 2 in this comment. As per our understanding, task abstraction was developed to identify similarities in the tasks that may seem like apples and oranges. To ensure that such task similarities do not go unnoticed, in Challenge 2 we argue that authors present an abstract task description along with the domain-specific task description. With this challenge, our primary motive was to foster a practice within the visualization community to promote abstraction of tasks. In C2: Guidelines for Researchers, we have mentioned and discussed the advantages of task abstraction. We have also expanded on the benefits and added another motivation that we believe will further enhance our push towards task abstraction. We consider a significant goal of a research article to be impact and outreach. Therefore, we think that adding task abstractions will allow authors to gather a larger audience for the paper.

This is all to say that the authors, if they really want me on board with their guidelines for researchers, needed to do a bit more rhetorical work (and really move this submission from

survey to position paper). If they are satisfied with reporting on their survey results, then I wouldn't necessarily bother.

---

## Reviewer 2

Overall, I like the paper and its motivation. However, I am a little puzzled about some of the challenges (C1, C2, and C3) as well as the guidelines provided for some of the challenges (C4). I feel like many of the challenges raised by the paper are somehow considered by the existing taxonomies.

**Meta-Response:** We thank the reviewer for pointing out their concerns about our paper. We believe that the reviewer may have misinterpreted our primary objective, or we did not explicitly communicate it sufficiently. Especially in C2, C3, and C4, the reviewer points out that the challenges have been discussed previously in existing taxonomies, and they can be resolved through the use of taxonomy. We are on the same page with the reviewer. The challenges that we discuss have been previously expressed in the form of discussion and opinions in many research articles (as discussed in Related Work of the paper), but they have never been systematically curated. Also, before our work, no evidence can validate if the challenges pose a threat to existing research methods and practices. We believe our work's primary contribution is the systematic curation of the problems and supporting challenges with evidence.

In the revised draft, we have strategically added words, sentences, or styling to ensure that we can minimize this confusion for other readers.

1. Introduction paragraph 2: We have italicized the concluding statement. The statement explains that task-based challenges have been discussed previously but they are not systematically curated in comparative studies.
2. Introduction (Contribution) and Conclusion: We have changed the phrasing to highlight that our contribution is the systematic curation of task-based challenges from previous visualization articles.
  - a. The introduction changed the phrasing to: "In this paper, we identify task-based challenges that exist in comparative studies through an analysis of existing visualization articles."
  - b. In conclusion, we added: "We identified four task-based challenges that can potentially affect the validity and usability of a comparative visualization study through an analysis of existing visualization literature."

### C1: Insufficient Justification of Task Source

I agree that it is important to discuss the source of the tasks we use in our study. However, I feel like describing the task source is getting more common these days. I have reviewed over 30

evaluation papers over the last 5 years and I barely remember a paper that did not describe the source of the task explicitly in their submissions. So, I am not sure how much this has to do with the set papers reviewed in this submission. Many of the papers reviewed in this submission were published more than 10 years ago. So, I somehow can understand why some of the older papers do not have such details, but I think it is less of a problem these days (in particular if we look at the publications at tier 1 venues).

**Response:** We have added a caveat in the discussion section to highlight that the year of publication may affect the evidence we have gathered in our survey. We have also proposed methods that we plan to use in the future to further validate the challenges. However, we would also like to mention that the problem of C1 highlights the lack of necessary guidelines to conduct a task analysis for comparative studies. We argue that data analysis, experiment design, infrastructure to use for evaluation studies have been better studied, than guidelines to choose and select tasks for an evaluation study. In the revised C1: Guidelines for Researchers, we discuss that the community needs to invest more effort into developing guidelines for selecting evaluation tasks.

#### C2: Missing or Incomplete Task Abstraction

I have a hard time understanding why existing taxonomies that we have don't perform the task abstraction. For example, the paper argues that we need to go through a process called "task abstraction". To the best of my knowledge, some of the existing taxonomies are already somehow doing task abstraction. For example, Taxonomy by Amar et al. first collects the raw questions/tasks. They then go through several steps to categorize these tasks into higher-level categories by removing unnecessary attribute names. Or Saket et al. (Group-Level Graph Visualization Taxonomy) collect the tasks by talking to experts. They then categorize the tasks into different groups where they also remove the domain language from the tasks.

I would like the authors to address this point. If what the existing taxonomies do is somehow similar to the task abstraction then why don't we tell people to use existing taxonomies? Why do we add another level of complexity which is introducing task abstraction to the end-users? However, if what the paper means by task abstraction is different from what the existing taxonomies do then I would like the paper to clarify/highlight that.

**Response:** As explained in response to the primary reviewer, we have added more information about the task abstraction to clarify our meaning and understanding of task abstraction. Further, we are not arguing for creating a new task taxonomy. Existing taxonomies have proved to be capable of task abstraction. In this paper, we identify that researchers need to adopt the published task taxonomies for task abstraction. We observed that some papers do not abstract their tasks and leave the work to the readers, which we argue is problematic and can be fixed easily if the authors present an abstraction of their task using an existing taxonomy.

### C3: Inconsistent Task Description Format

I am not sure how this is a problem if we encourage the authors to use existing taxonomies that are well-recognized and mention which taxonomy they used (C1). Even if other researchers don't understand the terminology used for tasks in a paper, they could always refer to the taxonomy used by that paper.

**Response:** We argue that the reviewers' comment regarding task description does not directly apply to C3. The goal of C3 is the development of a standard to report task descriptions in comparative studies. We do not believe this information is available in task taxonomies. Therefore we will remain with our challenge and argue that at the very least, researchers follow a TDAI format where in addition to the task description, authors present task abstraction and instruction.

### C4: Knowledge Gap in Task-Based Evaluations

The paper argues that existing comparative studies evaluate only a subset of the task design space. The paper then suggests ``To determine the task design space, a user should enumerate the possible combinations of analytical tasks a user can perform within an abstraction framework''. I am not sure if I understand why existing taxonomies are not doing this! Users can go through the taxonomies and pick representative tasks from different categories for their studies.

**Response:** We agree with the reviewer, and believe that the task abstraction frameworks implicitly provide task design space. We have modified the C4: Guidelines for Researchers to reflect the changes. Instead of asking the researcher to enumerate the task design space, we refer to them to use task design space presented in existing taxonomies and task abstraction framework.