Is This (Panel) Good Enough for IEEE VIS?

IEEE VIS 2022 Panel Proposal

Panelists: Cody Dunne[1], Alexander Lex[6], Torsten Möller[2], Alvitta Ottley[3], Melanie Tory[1]
Organizers: Robert S Laramee[4], Petra Isenberg[5], Tobias Isenberg[5]

[1]Northeastern University, USA, [2]University of Vienna, Austria, [3]Washington University, St Louis,
[4]Nottingham University, UK, [5]INRIA, France, [6]University of Utah, USA

INTRODUCTION

"The academic review process is broken," is a statement one often reads or hears. After getting the reviews back from the IEEE VIS conference, likely 75% (or so) of us agree. But is it really? The goal of this panel is to discuss the review process of the visualization community (broader than just IEEE VIS) and to brainstorm ways to improve upon it or to come to the conclusion that everything is fine. There are several concrete reasons why we believe this would be helpful:

A. It is difficult to find (qualified) reviewers.
What incentives could be put in place to accept reviewing duties? Or how can the reviewing load be lowered for our community?

B. Disagreement on acceptable contributions.
b1. Are NULL results acceptable, assuming the experiment was performed correctly?
b2. Is it acceptable to reject a paper on the grounds of "the contributions aren't strong enough for IEEE VIS?"
b3. Is X (e.g., literature reviews) an acceptable contribution for VIS papers and can and should reviewers and, in particular, PC members be held accountable to it?
b4. Should user study papers require institutional ethics approval and pre-registration or not?

C. Noise and randomness in what should be an objective assessment of research
In the light of the NeurIPS experiment of 2014 [1], that highlighted the noisiness of the review process, as well as the repetition in 2021 [2]:
c1: How noisy is the review process in our community and what consequences does this have, especially on early career researchers?
c2: Since NeurIPS hasn't found a way to address the aforementioned noisiness, is the review process truly broken or are there innovative ideas we can experiment with in the VIS community?
c3: Would a journal-first or journal-only process for submissions be able to address the noise issue, or would it cause other changes that we may not want to see (more papers to present, different type of papers to present, even more of a concentration on IEEE VIS as a venue, even further growth of the conference)?

D. Changes in the review processes of other venues

Is it worthwhile to think of an OpenReview system (reviewers remain anonymous, but reviews are part of something like https://openreview.net/), as is common in some other communities, in order to facilitate:

- *analysis*: open up the reviews for data analysis
- *transparency*: show the variety of the collected reviews
- *education*: help future authors to better understand criteria and culture of the VIS community.

E. The visualization field is growing
How do we deal with the growth of the field and conference in general and ensure that we remain sustainable?
e1: What can we learn from other fields which grow similarly or even more (e.g. NeurIPS)?
e2: More tracks, even less time for presentations, more concentration on full papers, …? Which of these strategies serves us as a community best?

Other questions to address may include:

- Can we assess the "strength" of visualization results?
- Can we define common thresholds (e.g., for literature reviews) etc.?
- The practicality about certain experiments, e.g., How about repeating "the NeurIPS experiment" at VIS?
- Can we collaborate between different venues (IEEE VIS, TVCG, EuroVis, PacificVis, CGF, CG&A, C&G, etc.) to reduce the community reviewing load?

The reason for asking all these questions is not only that we struggle with these points as authors, but also as seasoned reviewers. We believe these are difficult questions that don't have a simple answer.

For that reason it would be good to keep the conversation going and in the open and to crowdsource for solutions.

WHY THIS PANEL AT IEEE VIS 2022?

This is an important and timely theme for the visualization research community that addresses interesting, difficult and challenging questions. To the best of our knowledge, no such panel has ever been presented. This central topic touches on the experience and interest of every researcher in visualization. It is not only of interest to the VIS community if the lessons learned can be transferred to the other events in the future. It should be especially interesting for both experienced researchers and newcomers to the field. There is a wide variety of opinions on this topic within the visualization research community. We think a panel addressing the topic of contributions in visualization research will form the basis of lively discussions for the panel and more from the audience.

PANEL FORMAT AND LOGISTICS

The panelists will present their positions addressing each question posed in the introduction.
- The introductory remarks will be made by Bob Laramee. His introduction will last for 5 minutes.
- He will chair the panel and he himself is not a panelist.
- Each panelist will be given 10 minutes, for a total of 60 minutes of presentations.
- This will provide approximately 35–55 minutes of audience participation in the discussion.
- All panelists will have the opportunity to offer a summary view at the end of the panel (2 minutes each).

The panel chair will solicit audience feedback after the position statements have been delivered. The

panel format will also be described in the panel opening.

PANEL POSITION STATEMENTS

Cody Dunne

Many results presented at VIS are hard to reproduce, replicate, evaluate, or extend. I want to make it easier for us as a community to validate and build upon each other's work.

Several fields have suffered from a replication crisis, in which key findings that form the foundation of the field and can have life-changing impacts do not hold up under later scrutiny. VIS has fortunately avoided much of this problem—to our knowledge. Unfortunately, our community has not required authors to provide sufficient supplemental material and documentation for future researchers to actually test many of our findings.

VIS papers often omit key supplemental materials, including a study's stimuli, the datasets used in a benchmark, the choices made in statistical analysis code, or the source code of techniques and systems. Without these, future researchers are left to either (1) hope they can get the details by writing the original authors or (2) make their best guess. At minimum, omitting these materials results in substantial work for any researcher trying to reproduce what was presented.

When supplemental materials *are* shared, they are often placed on personal or group websites, GitHub, or other unreliable repositories. Websites disappear or are subject to link rot. GitHub may one day disappear like CodePlex or Google Code or have commercial restrictions placed upon it. We can avoid this problem by requiring authors to

upload materials to reliable and open archives with long-term survival plans. E.g., OSF.

My own work has run into extensive barriers put up by authors omitting key materials or hosting them on unreliable repositories. Graphs for case studies and benchmarks, once hosted on personal and group sites, have disappeared entirely. Writing to authors for data and code often gets no reply, or, if there is one, that materials cannot be shared because they are lost, in a poor state, or company secrets. (This, embarrassingly, includes some of my own research.) Sometimes we found other datasets to use. Other times we tried re-implementing software and gave up due to time or poor specification. In some cases, we chose to compare against fewer or worse techniques that were easily available to avoid reinventing the wheel.

I argue that the visualization community should require all supplemental material necessary to recreate the work be included with each submission—for reviewers to evaluate and future readers to use. Further, materials should be hosted on free, open, and reliable archives. Naturally, there are situations where sharing is impossible or unethical. Our community should develop expectations for what are valid and invalid justifications for omitting materials, and provide these for reviewers to use while determining whether the paper is truly making a contribution to the community.

Alexander Lex
B3 - Contributions
VIS/TVCG needs to break out of the "one type of research paper" format. While the conference does have all kinds of formats, the archival output is very monolithic. I would argue that we need different formats published in our journals. For example,

there is no culture of written discussion and responses (letters to the editor). We should try to move academic debate, such as discussing issues with papers away from social media and replace it with well reasoned, calmly written letters to the editor and responses. Maybe there could also be a lower barrier process, such as with a moderated blog or moderated post publication reviews. Other journals also have dedicated formats for surveys and application notes (e.g., Oxford Bioinformatics) which could be adopted at VIS.

D. Review Process.

The review process (with respect to the quality of reviews) at VIS is working fairly well compared to other fields. On average, reviews are well reasoned and thoughtful. However, I believe that more transparency in the review process would be beneficial. This could include publishing the submitted draft manuscript, the reviews, and subsequent versions. I suspect this would lead to higher quality submissions and reviews. I would argue against publishing reviewer identities, since this can lead to animosity and make junior reviewers vulnerable.

VIS needs to move to a **mandatory double blind** submission system. There's ample evidence that double blind reviewing results in more just review decisions (see references at http://double-blind.org/). I haven't heard convincing arguments against double blind reviewing. Conflicts of interest should be managed on a reviewing system level, and given that the overwhelming majority of CS conferences use double blind reviewing (see link above), and that VIS already uses optionally double blind submissions it can't be an insurmountable problem. I would also argue that anonymity should preserved also for at least the secondary program committee member, who currently can see the authors even of papers submitted in double blind format.

B4. Ethics Guidelines

VIS should have enforceable ethics guidelines (i.e., it should be possible to reject a paper based on ethics violations). IRB approval is not a good standard because they're governed by local laws and some organizations don't have IRBs. However, the community could require some general guidelines that make authors demonstrate that they have considered human subjects' concerns, including but not limited to fair pay for study participants. Papers should also discuss ethical implications more broadly. For example, research in support of cryptocurrencies would need to provide an analysis of potential harm and benefits.

C3. Journal-First

I would advocate to *not* change VIS to a different submission model, e.g., with a quarterly journal deadline. I personally like having a deadline; deadlines are liberating in that they allow us to focus. With a deadline the *important* work of doing research is also the *urgent* task of the day. Otherwise, not-important but urgent tasks tend to get in the way of focused work. However, I realize that deadlines aren't great in other ways, and that other people might be better at prioritizing important over urgent work. But I do think that the ability to present a TVCG paper at VIS basically solves that problem. We should also aim to strengthen alternative venues, such as EuroVis, so that we maintain a diversity of high quality options to disseminate our research.

Torsten Möller

In light of many years of anecdotal evidence (as a reviewer), we know that the reviewing process is noisy. However, the NeurIPS experiments backed up this claim with hard evidence by showing that "about half the papers published at the conference would have been different" [1]. Most people didn't think that it is *that* noisy. If we would apply such a system for determining who gets welfare assistance and who doesn't, there would be a public outcry on all major networks, no matter what country. Hence, I would claim, that the process is flawed. But beyond this, it impacts the career of PhD students (by lengthening their study time in a low-paying job or even changes their career path), tenure-track professors (by jeopardizing their tenure chances) and hence is not unjust, but also unethical. Therefore, it would be good to collectively brainstorm on how to change the system. At the same time, one has to acknowledge that it is not all bad -- it perhaps works for a large part of the community, but this is still not an excuse to ignore this problem for the vulnerable members of our community.

One of the suggestions resulting from the NeurIPS experiment was "the real conclusion of the experiment is that the community should place less onus on the notion of 'top-tier conference publications' when assessing the quality of individual researchers." [1] Hence, I would like to propose to not focus on 'accept' vs. 'reject' but to try and shepard all papers with a good idea to a successful publication in one of the venues. A simple approach would be to be very restrictive on the "reject" option and instead open it up to "major revision with changes most likely appropriate to a specific venue including, but not limited to IEEE VIS, EuroVis, PacificVis, etc. This might help with the reviewing fatigue. The obstacles to overcome would be:

a) elitism of the reviewers
b) the focus on acceptance rates as the sole evaluation criteria for a venue
c) the collaboration across publishers (IEEE, ACM, EuroVis, etc)
However, this only makes sense, if the review process at VIS is as noisy as the review process at NeurIPS. Hence, as a first step it would be necessary to repeat the NeurIPS experiment at VIS.

Alvitta Ottley

In research, it is just as important to mark known dead ends and dangerous spots as blazing the paths that people can travel safely. The undervaluation of null results impedes research progress.

About ten years ago, I ran a user study investigating the impact of visualization on statistical reasoning, with the hypothesis that visual representations would improve reasoning accuracy. Surprisingly, our results showed no significant difference between the tested text-only and visualization conditions. Was something wrong with the experiment design, execution, or data analysis? Why didn't visualization *help* despite prior evidence that it should? This null result rendered the research project unpublishable, or so we thought. That same year, another group of researchers published similar work with virtually identical findings but with an obligatory follow-up study showing that visualizations can help under particular conditions (if we remove the numbers from the text) [3]. Seeing this other paper validated my user study findings, so I pulled my project from the figurative filing cabinet I had stuffed it in and submitted a tech report at my university [4]. [The project was again unpublishable since a replication study is often a kiss of death for publication, but let's focus on one publication bias at a time.]

Today, null results remain rarely published, even though their contributions can be the catalyst for us to reexamine what we think we know about visualization. Further, the availability of null findings

would save time and resources from revisiting research questions already answered but are not readily accessible to other scholars. This publication bias exists beyond the visualization community, with similar declarations of selective reporting throughout the social sciences. For example, one meta-analysis from the social sciences found that "strong results are 40 percentage points more likely to be published than are null results and 60 percentage points more likely to be written up" [5]. I argue, however, that the impact of selective reporting is more significant in a relatively young and small community such as VIS. It limits our collective understanding of the state of our knowledge and impedes our ability to carve out paths for future discoveries.

So, what now? How do we fight an invisible enemy since we can't see what's not published? The primary conclusion from the meta review mentioned above is that "authors do not write up and submit null findings." This tendency to file away null results is probably also common in the VIS community. Many authors do not write up their findings because such papers seldom make it through the review process. Anecdotally, I still need to fight an internal battle to resist the urge to set aside my findings that do not show solid statistical results, even when the hypothesis reflects long-standing dogma and the null findings serve as a crucial challenge to the status quo. This is because my experiences as an author of papers with null findings and as a reviewer indicate that the proverbial Reviewer 2 will likely say something to the effect of "the contributions are not strong enough" or "the project is not mature enough for publication."

This type of feedback is deeply misguided. Judging a contribution based on whether statistical test results fall above or below $\alpha = .05$ misses the big-picture contribution. In research, it is just as important to mark known dead ends and dangerous spots as blazing the paths that people can travel safely. Further, evaluating a paper based on an arbitrary cut-off may encourage "p-hacking" – selectively reporting results showing statistical

significance and contribute to the replication crisis. VIS must seriously assess how we review research products and how this process might spawn harmful research practices. We can begin by ending the culture of devaluing null results and developing guidelines for evaluating them.

Melanie Tory
"Diversity is being invited to the party; inclusion is being asked to dance." [Verna Myers]
I want VIS to be a place where everyone feels invited to dance. Reviews should be an enabler, not a barrier.

Fostering diversity and inclusion in the field of visualization is critical to our long-term impact. It is well established that diverse teams develop more creative solutions. While all forms of diversity are important, here I focus on intellectual diversity, encouraging our field to be inclusive to all disciplinary backgrounds and research approaches. The interdisciplinary nature of visualization, often described as both an art and a science, creates a unique opportunity to bring together a rich collection of ideas and approaches that can help us think in new and exciting ways. Yet we tend to fence ourselves in with rules, institutionalized both in paper submission guidelines and in reviewers' minds, that may make sense for mainstream approaches but can inadvertently exclude others.

Should VIS be restricted to specific contribution or paper types? Should user studies require institutional ethics approval and pre-registration? Should reviewers be required to have X-specific experience? My answer to these sorts of questions is no. While such questions are usually asked with the intention to ensure "good science", they erect barriers to intellectual diversity. For instance, while paper and contribution "type" exemplars have value for helping new students grasp the diversity of ways a paper can be crafted, limiting ourselves to a

defined set may discourage unexpected ideas that we haven't imagined yet. The very notion that we have such defined types may make those contributions that don't "fit the mold" feel unwelcome, or lead reviewers to reject otherwise interesting contributions. Similarly, mandatory human research ethics approval excludes organizations without ethics boards (e.g., industry). And while mandatory preregistration may guard against p-hacking in quantitative research, it excludes virtually all qualitative methodologies, for which preregistration is at a minimum meaningless and in the worst case a violation of good scientific practice. The question we should be asking ourselves when we review a paper is not whether it meets X, Y, and Z arbitrary criteria, but rather whether it contributes a new idea of value to VIS or stretches our thinking with new and interesting perspectives.

PANEL BIOGRAPHIES

Cody Dunne

Cody Dunne is an Assistant Professor at Northeastern University. His research focuses on helping people explore and understand complex data—in particular data that combines aspects of network topology, position in space, values of attributes, changes to all of these over time, and how changes or events can happen in sequence. Prior to joining Northeastern, Cody was a research scientist at IBM. Cody received his PhD in computer science under Ben Shneiderman at the University of Maryland Human-Computer Interaction Lab in 2013 and earned a B.A. in computer science and mathematics from Cornell College in 2007. Cody is currently serving on the IEEE VIS Open Practices committee and has a strong history of publishing supplemental materials with his papers.

Alexander Lex

Alex is an Associate Professor of Computer Science at the Scientific Computing and Imaging Institute and the School of Computing at the University of Utah. He directs the Visualization Design Lab where we develop visualization methods and systems to help solve today's scientific problems.

Before joining the University of Utah, he was a lecturer and post-doctoral visualization researcher at Harvard University. He received his PhD, master's, and undergraduate degrees from Graz University of Technology. In 2011 he was a visiting researcher at Harvard Medical School.

He is the recipient of an NSF CAREER award and multiple best paper awards or honorable mentions at IEEE VIS, ACM CHI, and other conferences. He also received a best dissertation award from my alma mater. He co-founded datavisyn, a startup company developing visual analytics solutions for the pharmaceutical industry. Alex has been involved with IEEE VIS in many capacities; most recently he served as an Area Papers Chair (2021) and in the reVISe committee.

Torsten Möller

Torsten Möller is a professor of computer science at the University of Vienna, Austria, since 2013. Between 1999 and 2012 he served as a Computing Science faculty member at Simon Fraser University, Canada. He received his PhD in Computer and Information Science from Ohio State University in 1999 and a Vordiplom (BSc) in mathematical computer science from Humboldt University of Berlin, Germany. He is a senior member of IEEE and ACM, and a member of Eurographics. His research interests include algorithms and tools for analyzing and displaying data with principles rooted in computer graphics, human-computer interaction, signal processing, data science, and visualization. Since 2018, he serves as the editor-in-chief for IEEE Computer Graphics and Applications. He was appointed Vice Chair for Publications of the IEEE Visualization and Graphics Technical Committee (VGTC) between 2003 and 2012. He has served on a number of program committees and has been papers co-chair for IEEE Visualization, EuroVis, Graphics Interface, and the Workshop on Volume Graphics.

Alvitta Ottley

Alvitta Ottley is an Assistant Professor in the Department of Computer Science and Engineering at Washington University in St. Louis. She also holds a courtesy appointment in the Psychological and Brain Sciences Department. She directs the Visual Data Analysis Group, where she evaluates and designs visualization techniques for helping people explore, reason, and make judgments with data. Her research applies machine learning and artificial intelligence to automatically learn goals, cognitive traits, and future behavior from interaction data with visualization. Dr. Ottley received her Ph.D. from Tufts University and is a recipient of the NSF

CRII and CAREER awards. Dr. Ottley has served on the IEEE VIS and ACM CHI program committees and the IEEE VIS, ACM IUI, and CMD-IT/ACM Tapia organizing committees. In addition, she has organized the Visualization for Communication Workshop at IEEE VIS and has been the Papers Co-Chair for the IEEE Symposium on Visualization in Data Science from 2020 to the present.

Melanie Tory

Melanie Tory is Director of Data Visualization Research at the Roux Institute, Northeastern University. Her team focuses on human-data interaction for health and engineering applications, and the interplay between visualization and AI. Previously at Tableau, Melanie managed an applied user research team and conducted research in natural language interaction with visualizations. As an Associate Professor at the University of Victoria, she explored topics such as collaborative visualization and personal visual analytics. Melanie earned her PhD in Computer Science from Simon Fraser University and her BSc from the University of British Columbia. She is on the IEEE VIS steering committee and serves as Associate Editor of IEEE Computer Graphics and Applications, IEEE Transactions on Visualization & Computer Graphics, and Computer Graphics Forum.

REFERENCES

[1] Corina Cortez and Neil D Lawrence, Inconsistency in Conference Peer Review: Revisiting the 2014 NeurIPS Experiment, Technical Report, 2021 https://arxiv.org/abs/2109.09774 (last access July 2022)

[2] Alina Beygelzimer, Yann Dauphin, Percy Liang, and Jennifer Wortman Vaughan; NeurIPS 2021 Program Chairs, The NeurIPS 2021 Consistency Experiment. Blog post, https://blog.neurips.cc/2021/12/08/the-neurips-2021-consistency-experiment/ (last access July 2022)

[3] Micallef, L., Dragicevic, P., & Fekete, J. D. (2012). Assessing the effect of visualizations on Bayesian reasoning through crowdsourcing. IEEE transactions on visualization and computer graphics, 18(12), 2536-2545.

[4] Ottley, A., Metevier, B., Han, P. K., & Chang, R. (2012). Visually communicating Bayesian statistics to laypersons. In Technical Report. Tufts University.

[5] Kepes, S., Banks, G. C., & Oh, I. S. (2014). Avoiding bias in publication bias research: The value of "null" findings. Journal of Business and Psychology, 29(2), 183-203.