Pipelines Bent, Pipelines Broken: Interdisciplinary Self-Reflection on the Impact of COVID-19 on Current and Future Research (Position Paper)

Priscilla Balestrucci*
Applied Cognitive Psychology
Ulm University, Germany

Katrin Angerbauer†
Visualization Research Center (VISUS)
University of Stuttgart, Germany

Cristina Morariu
Visualization Research Center (VISUS)
University of Stuttgart, Germany

Robin Welsch
Human-Centered Ubiquitous Computing
LMU Munich, Germany

Lewis L. Chuang
Human-Centered Ubiquitous Computing
LMU Munich, Germany

Daniel Weiskopf
Visualization Research Center (VISUS)
University of Stuttgart, Germany

Marc O. Ernst
Applied Cognitive Psychology
Ulm University, Germany

Michael Sedlmair
Visualization Research Center (VISUS)
University of Stuttgart, Germany

Abstract

Among the many changes brought about by the COVID-19 pandemic, one of the most pressing for scientific research concerns user testing. For the researchers who conduct studies with human participants, the requirements for social distancing have created a need for reflecting on methodologies that previously seemed relatively straightforward. It has become clear from the emerging literature on the topic and from first-hand experiences of researchers that the restrictions due to the pandemic affect every aspect of the research pipeline. The current paper offers an initial reflection on user-based research, drawing on the authors' own experiences and on the results of a survey that was conducted among researchers in different disciplines, primarily psychology, human-computer interaction (HCI), and visualization communities. While this sampling of researchers is by no means comprehensive, the multi-disciplinary approach and the consideration of different aspects of the research pipeline allow us to examine current and future challenges for user-based research. Through an exploration of these issues, this paper also invites others in the VIS—as well as in the wider—research community, to reflect on and discuss the ways in which the current crisis might also present new and previously unexplored opportunities.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

The outbreak of the COVID-19 pandemic has had a profound impact on every aspect of society. Research centers and academic institutions were not exempt from the disruption brought about by the crisis, and scientists in all fields and disciplines have had to restructure their entire work pipeline almost overnight in order to adapt to this unprecedented situation [37].

Empirical evidence on the impact and long-term effects remains necessarily fragmented and still incomplete, as the situation continues to develop in unpredictable ways. It has nonetheless become apparent that the pandemic will have long-lasting implications for research practices as well as for the life and career of researchers [39, 51].

Due to the predominantly social character of restrictions implemented globally to combat COVID-19, many issues have emerged for conducting scientific research, ranging from general problems concerning social interaction to more specific complications related to the nature of experimental research. For example, recent surveys have addressed the issues associated with the closure of research facilities and the subsequent need to work off-site, primarily at home. This radical shift in working habits has given rise to various problems, from mounting psychological stress on researchers to more systematic issues related to managing childcare and gender equality [2, 37]. Due to restricted access to typical facilities, not only have researchers had to interrupt current projects, but the planning and implementation of future research have become more difficult, as such projects would need to consider elements such as social distancing, hygiene, and other concerns. These requirements will have an evident impact not only on methodological aspects of future studies, but the consequences of these changes will likely be reflected in the ways in which research studies will be funded and published in the future [13].

Lastly, normal practices in the communication and dissemination of scientific research have been heavily reconfigured. The vast majority of scientific conferences have been either postponed or outright canceled, or else they have been moved online [1]. Since these events not only allow for the communication of research but also facilitate interaction within the community, this change will likely affect the way in which researchers can maintain and develop connections with other members in their field. From an analysis of the emerging literature on the effects of the pandemic, it is clear that aspects of the entire pipeline of the research workflow, both in the short and in the long term, have been affected by the current crisis (Figure 1). What is not immediately apparent from the existing literature, however, is the wider cumulative effect of the crisis, and the ways in which such issues should be addressed on broader structural level. In other words, these various social and methodological aspects must be considered together as an aggregate when discussing possible future scenarios for the research communities.

While the pandemic has had a demonstrable impact on every field of research, there have also been consequences that are specific to human-based research. More than in any other fields, methodological and social aspects are intrinsically related in studies that rely on human participation. As a consequence, it has become even more important to consider the whole research pipeline in order to determine how the pandemic will affect the work done by these research communities. The need to interact directly with participants, both online and in person, poses unique challenges, which are shared by researchers across all such research disciplines. The complex situation of user-based studies thus requires particular consideration on methodologies and common practices that previously

*e-mail:priscilla.balestrucci@uni-ulm.de
†e-mail: Katrin.Angerbauer@visus.uni-stuttgart.de

Figure 1: The research pipeline.

seemed relatively straightforward. While all aspects of the research pipeline must clearly be reconsidered in light of the current situation, especially when dealing with user studies, in this paper we argue that it is also possible—and necessary—to reevaluate the disruption of research activities as an occasion to explore and develop new opportunities for the field.

The main contribution of this paper is a reflection on the current and ever-evolving situation of user-based research following the outbreak of the COVID-19 pandemic. Toward that goal, and to provide support to the presented position, we conducted a preliminary survey with 29 participants, collecting potential issues, use cases, and strategies for user-oriented research. We use these results, together with our own experiences in an interdisciplinary team of visualization, HCI, and perception researchers, to reflect on which aspects of the research pipeline are affected by the changes, both in the short and long term. In so doing, we hope to provide comments and ideas on the ways in which these changes can represent challenges but also opportunities for both the wider research community, and visualization researchers.

2 Background and Related Work

We first review work related to user-oriented methodologies, and assess the impact of COVID-19 on research and evaluation pipelines in the different research communities. We then assess pipelines and evaluation methods commonly used in visualization, and assess them in terms of their vulnerability to the impact of COVID-19.

2.1 Related Work

Direct human participation is essential not only for research in the VIS community but also in several other fields, such as experimental psychology, behavioral neuroscience, and human-computer and human-robot interaction [14,40]. While the more general challenges faced by researchers also exist in these fields, the specific requirements for conducting human-based research might differ based on the unique characteristics and challenges in a research field.

With the shutting down of research facilities, many researchers have shifted the focus of their work to make use of online or remote resources for their studies [42]. In some cases, the shift to in-person to online or remote studies has been made possible either through the adaptation of the experimental questions or through the use of customized tools [29]. In many cases, however, it has proven to be impossible to make this shift either due to the lack of available technology or to the nature of the originally envisioned study [14].

Of specific interest here is the working paper by the HCI researchers Schmidt and Alt [43]. They offer a set of evaluation approaches that might offer an alternative to lab studies in HCI. Ideas such as “Using Existing Data Sets”, “Engage With Users Through Remote Communication”, and “Appropriate Your Research Question And Method To Users Who Are At Home And Where You Have Not Direct Access” are likely also interesting for visualization researchers.

At the time of writing this article, some research facilities are reopening and adopting different guidelines depending on location and the decisions of individual institutions. As it is in some cases possible to run user studies again, it is clear that the safety of the participants remains the priority. Given the nature of this particular context, recommencing user studies raises not only practical challenges but also ethical concerns [53].

2.2 Research Pipelines and Evaluation Methods in Visualization

Visualization research follows several different pipelines, which do not fit into a one-size-fits-all box. These different ways of doing research can be seen in the five typical categories for papers: technique, evaluation, system, application/design study, and theory [35]. Some specific pipeline models to guide visualization research in these categories do exist, such as the nine-stage model for design studies [45] and the processes for evaluating visualizations by Carpendale [10]. Since visualization represents an intrinsically interdisciplinary domain, work in the field is also strongly shaped by research pipelines from other areas, such as HCI [27] and perceptual psychology [15,52].

Doing research is inherently coupled to how contributions are evaluated. This topic has been the core focus of the BELIV community for the last 14 years, and many papers have been published on the topic by now. The systematic literature analyses by Lam et al. [26] and Isenberg et al. [21] group evaluation practices into seven or eight scenarios, respectively. The evaluation scenarios of ‘algorithmic performance analysis’ and ‘qualitative results inspection’ should not be substantially affected by COVID-19 as they do not involve humans in the evaluation process. All other scenarios, however, are intrinsically tied to some form of user testing and might thus be strongly affected by the pandemic. Replacing methods by other methods, however, is also not a trivial task, as the choice of evaluation methods usually is dictated by the research question and contribution type targeted [32,36].

Here, our emphasis is on the empirical research methods in visualization that involve user testing. These methods are specifically vulnerable to a lockdown situation. In particular, controlled lab studies seem to fall into this category, such as perceptual studies or studies of visualization and interaction techniques. Other research pipelines might be less affected. Design studies with qualitative pair analytics [3] and case studies [48] could potentially be shifted to online interviews and collaborations without substantial hurdles. Observational studies at the actual workplaces of users [21,26] might be extremely difficult though. Online studies, such as on Mechanical Turk [18] may be hardly affected, and could even be a good surrogate for planned lab studies. While a study in an online setting is less controlled than a laboratory one, the validity of results could be reinforced by a larger number of users, which would otherwise be impossible to recruit on a university campus. Heer and Bostock [18] offer recommendations for such crowdsourcing studies and make the case for a more reproducible setup.

Controlled experiments in visualization follow the typical steps that are also common in other areas such as HCI and applied psychology: study design and implementation, data analysis, and reporting the results [10]. This basic pipeline is enriched with steps such as acquiring ethics approval, pre-registration of the study, properly accounting for General Data Protection Regulation (GDPR) aspects, and making the data publicly available—steps that have also been discussed at past BELIV workshops [11,17,25].

While typical visualization studies are conducted with a computer screen and mouse and keyboard, more sophisticated hardware setups are also becoming more common in empirical visualization research. Studying collaboration of multiple users on an interactive tabletop [20] might, for instance, be hardly possible under COVID-19 distancing rules. Similarly, evaluation setups using eye trackers [7, 50] might get completely removed from the equation due to hygiene requirements. Work that seeks to simulate eye movements without the need of special hardware, such as BubbleView [23] and Fauxvea [16] might become more relevant. These approaches aim to blur the screen with the exception of a small circular area, which the user can shift to reveal the screen content or where the authors aim to predict saliency maps of user attention.

3 Survey

In this section, we now present the online survey we conducted with 29 researchers. After describing the survey layout, we will present the results. The discussion of the results will follow in the subsequent section.

3.1 Methods

To gather an impression on research experiences, we conducted a short survey on the status of user studies and, in particular, how they were being affected by COVID-19. A full version of the survey is available in the supplementary material.

First, we asked researchers which of the following four scenarios better described the state of their ongoing study:

- *Case 1: A remote or online study was planned prior to the lockdown.*
- *Case 2: An in-person study was originally planned, but was switched to an online study.*
- *Case 3: An in-person study was planned, but it was not switched to online / remote (i.e., it was paused the or continued as planned.).*
- *Case 4: A study was planned to tackle issues related specifically to COVID-19.*

We then asked for more detailed information on the reported study in order to better understand which aspects along the research pipeline were affected by the outbreak of the COVID-19 pandemic. Some questions were common for all cases and concerned the current state of the study under discussion. Other questions instead related to the specific context, i.e. the type of case that was being reported. For questions concerning the state of a study, i.e., the 'stage' of the pipeline at the time of the survey, participants were asked to answer by selecting one of multiple choices. For questions that were more specific to the case under consideration, participants were asked to answer more open-ended questions or to select one or more choices from a given list.

The survey was created using Google Forms and circulated via social media and through the network of the authors during April to June 2020. All participants responded via the online form, although they were given the choice of leaving their personal information in case they wanted to be contacted for further clarifications. No personal data were collected unless explicitly agreed upon by participants.

3.2 Results

We collected 35 reports of study cases from 29 peer researchers mostly in the fields of psychology and computer science (namely, HCI and visualization researchers), see Figure 2. Of the 35 cases reported (Figure 3a), 6 had been planned as online or remote studies regardless of the pandemic (17%), whereas 10 studies were converted due to the change in the situation (29%). Several studies (16 out of 35, 46%) that had been planned as in-person studies were not converted to online or remote studies. Lastly, 3 studies (∼9% of our sample) reported tackling COVID-19-related issues. For studies that were not moved online, their situation had various outcomes (Figure 3b).

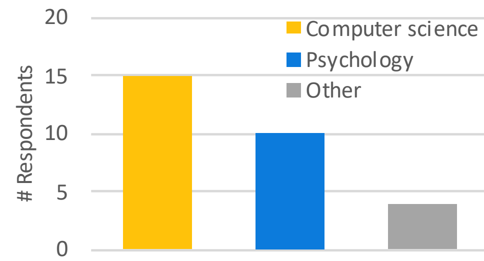


Figure 2: Respondents to the survey.

Figure 4a reports the heterogeneity of the current state of the studies conducted online or in remote form. For studies that were originally designed to be run in person, there were several changes that needed to be implemented, regardless of whether they were continued in person or moved online (Figure 4b).

While the governmental response to COVID-19 has been heterogeneous from country to country, most knowledge workers switched to home-office setups. This is also the case for the respondents of our study. While many people are now returning to work in the office across the globe, our survey was mostly concerned with studies run at the very beginning of the pandemic when this was not yet the case.

In the next section, we will interpret these results, connect them with our own experiences, and further discuss the qualitative feedback we got from the survey.

4 Discussion

In this section, we interpret and discuss the results of our survey, enriched with our own experiences and opinions. This part is mostly meant as an initial **position statement** and not a fully-fledged and evaluated research contribution. Also, we do not restrict ourselves to aspects that are strictly specific to visualization research, but also include aspects that might have an impact on a much broader scale. Our goal is to start a discussion and possibly inspire others on potential issues, solutions, and even opportunities. While our focus lies primarily on user-oriented research, many of the concepts under discussion can also be applied to other fields.

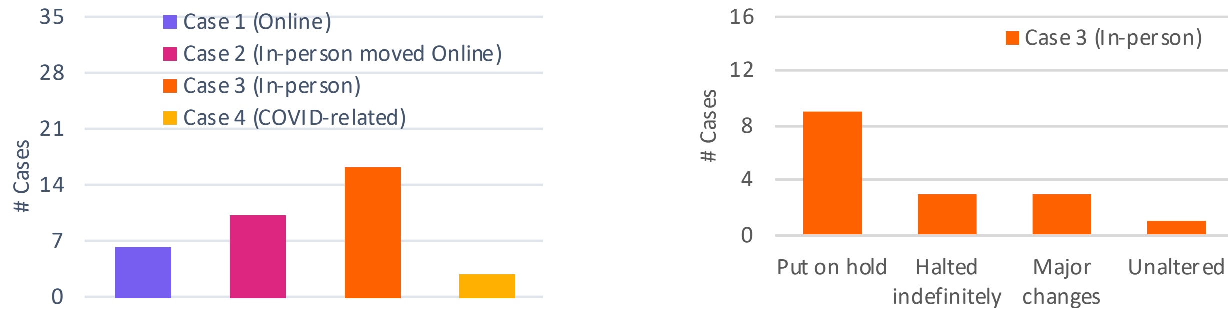
To this end, we organize our discussion along a typical pipeline for empirical research in visualization, consisting of the following steps:

- Conceiving research and research environment
- Study design
- Data collection
- Communication of results

4.1 Conceiving Research and Research Environment

The conceptual basis for research is the first step toward a successful research project and builds the most relevant framework for it. The discussion of the impact of COVID-19 on user studies often ignores this aspect. However, we found some indications in our survey that this stage of research is also affected.

In this sense, COVID-19 has a potential impact on research directions, or at least, on the details of a research question. Research ideas might need to be modified in order to continue user testing despite the lockdown of research facilities. This is especially critical



(a) Total cases reported

(b) State of studies in Case 3 (In-person)

Figure 3: Case studies in the survey.

in cases where a study had already passed the design phase, and was undergoing data collection when the lockdown measures were taken. In the most extreme case, some research concepts were no longer feasible, to the point that ongoing studies had to be put on hold or even be halted indefinitely. For example, one respondent noted: *"We cannot easily move this study online as it requires special equipment that cannot be replicated online."*

At the same time, completely new research questions might be triggered by the pandemic, as can be noticed by the fact that some of the participants to our survey reported planning or collecting data on user studies tackling COVID-19-related issues. Several visualization researchers have already worked on visualization systems for coronavirus-related data. Yet, we also see the opportunity for less immediate influences on research agendas.

Aspects of the research framework that we believe require further consideration are the funding situation, the legal and regulatory environment, and the collaborative nature of research work. Legal systems, policies, and regulations vary between countries, which necessitates reflections for international collaborations and joint research projects. For example, data privacy regulations and protection as well as ethics approvals differ widely between different regions of the world. These issues might become even more pronounced when studies are moved online. Some of the issues might also become extremely hard to address due to a lack of law expertise. Another concern for joint collaborative work relates to travel restrictions, since project meetings might no longer be feasible.

The above issues play an important role for running projects, but they could also impact the planning of, and application for, research grants. For example, there might be a trend toward more local or less collaborative projects, which would have a negative long-term effect on the research community.

4.2 Study Design

Re-designing user studies. The design of many user studies has become unsuitable for data collection during the COVID-19 crisis. Researchers reported that they were either forced to decide whether to run their study unchanged, develop an elaborate hygiene plan to decrease the risk of infection, change the study design to minimize risks, or stop the study altogether. Our survey revealed that the majority of studies (25 of 35; 72%) needed to be fully or partly redesigned. Some respondents (10 of 35; 29%) were able to move their studies online by changing the experimental design. Changes required for moving online included simplifying the instructions, changing target samples, and presenting fewer experimental trials. This move to online studies was typically accompanied by concerns regarding the level of stimulus control and technical affordances.

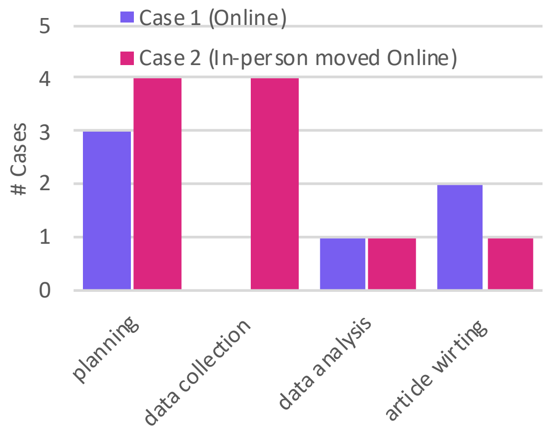
In-person studies. Studies that continued despite the crisis were required to employ strict hygiene protocols, which grossly modified the study design. This change did not necessarily work out smoothly. One respondent, for example, reported losing the possibility of interacting with a population that was undergoing a longitudinal study: *"[We might] discard the study because it is on development and the kids might have grown out of the interesting age by the time we can restart."* Another described a need to limit the number of sessions in a psychophysics experiment. This constraint led to a delay in data collection and ultimately forced the respondent to submit the study to another conference.

Study designs necessarily in conflict with social distance requirements or hygiene protocols were most heavily affected by the crisis. These were primarily multi-user studies in which participants needed to be physically present, such as studies relying on perception or using head/body-mounted equipment. For example, one respondent noted: *"The biggest uncertainty derives from collecting kinematics and EMG [electromyography]. How can you place electrodes on the participant body with social distancing?"* Most of these studies have been either fully canceled or postponed until the lockdown measures are removed or changed. One study was halted prematurely due to the crisis, however the researchers analyzed the data they had already acquired. It is noteworthy that research communities are currently developing new data collection protocols with an emphasis on hygienic practices (e.g. [49]), which arguably ought to have been considered even before COVID-19.

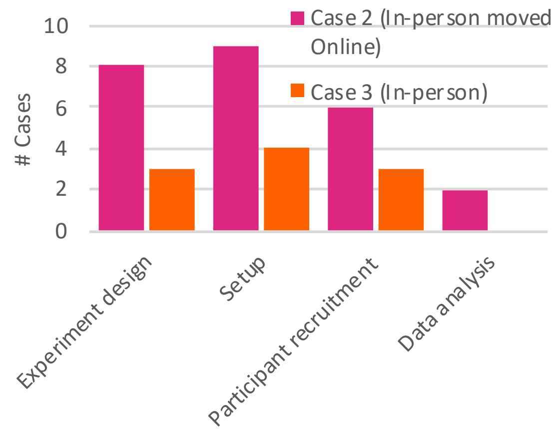
From in-person to online studies. Prior literature suggests that it is possible to move a study to a remote context by providing subjects with hardware. Mottelson and Hornbaek [34] could demonstrate that valid behavioral data can be obtained when using distributed remote testing in a VR-context. They showed that using a cardboard VR setup, user behavior for pointing, 3D tracing, and body ownership illusions was comparable to studies using a HTC VIVE inside the laboratory. As an additional advantage, such remote studies could target a large and diverse sample of users. Thus, moving from in-lab to out-lab studies may even beneficial for the validity of the conclusions drawn from the data. However, this scenario is only possible if the study design allows for mobile equipment. Stationary, highly calibrated, and expensive experimental setups, such as an eye-tracker, cannot make use of this approach.

It has to be noted that, in order to move a study online, experimental design, setup, and infrastructure had to be changed, which may also lead to less control over the experiment. One respondent noted: *"Less control about data presentation and collection devices and environmental circumstances during testing (time of day, distractions, ...)".*

Lastly, we recognize that moving a study outside the lab and onto



(a) Current state of studies in Case 1 (Online) and 2 (In-person moved online)



(b) Most problematic aspects for continued studies in Case 2 (In-person moved online) and 3 (In-person not moved online)

Figure 4: Current state of studies.

a large variety of screens, remote working set-ups bring a new set of challenges. User attention and perception might be affected by outside conditions, making it more difficult to measure the effect studied in the experiment. This is the reason why a careful planning and consideration of the possible confounds must be addressed during the design phase of the study. The negative effects resulting from moving a study outside of a laboratory setting can—at least in part—be compensated by employing a larger set of participants. While this can ultimately result in much higher costs of the study, it can also lead to more ecologically valid results [18].

4.3 Data Collection

In-person studies. For respondents who described continuing an in-person study, major changes to the setup were required, specifically concerning the implementation of a hygiene protocol. One respondent wrote, for instance, *"I'll have to limit the number of sessions per day, maybe to one or two, and disinfect all the equipment used, apart from including safety equipment for the examiners."* For laboratory experiments put on hold during the lock down, in general the research timeline was negatively affected. One researcher in our survey even spoke of an indefinite interruption for an experiment due to a change in professional position: *"I had 4 more participants to collect, which I couldn't because of the lockdown. I think I will never collect them as I am transitioning to a different position."*

Online studies. For crowd-sourced setups, there appears to be no noticeable difference so far in terms of the subject pool available. Demographic data showing who used the platform prior to the pandemic compares favorably with data from today [30, 33]. Yet, we could also reasonably assume that the participant pool could change in the long term, with participants becoming less diverse [31]. As highlighted in an article by the New York Times [44], it is unclear whether people who are unable to pay the bills are able to keep an internet connection or be willing to participate in research. Although most studies on Amazon's Mechanical Turk are paid, the average worker prefers steadier sources of income than research participation, such as image annotation or data collection tasks [30]. Based on the available information, it is still unclear what impact the pandemic will have on online user studies in the long term. For example, while several works addressed the issue of gender disparities in the working conditions following the pandemic [2, 37], to our knowledge there are no studies that focus on the change in participation of women in online studies.

Participant recruitment. Based on reports in the survey, participant recruitment was also heavily affected by the pandemic. For in-person studies, recruitment became either incredibly difficult or outright impossible. The situation further changed the participant sample. One respondent commented, *"We have originally planned to study performance of healthy sighted controls, to be compared with that of visually impaired participants. We are now running the study on healthy participants only, since we have lost access to patients."* For remote and online studies, one respondent actually reported the opposite effect, describing a benefit from the lockdown situation: *"We had planned a small expert study involving our colleagues and university network before releasing the main one on Amazon Mechanical Turk. We expected 10 participants, and got 55 responses instead. As a result, we decided a crowd-sourcing study was unnecessary, and the current results would be higher quality."* The participant further speculated that people involved had more available time to participate in studies online, at the beginning of the pandemic where many other job responsibilities were paused.

4.4 Communication of Results

A central aspect of the scientific pipeline is the communication of new scientific findings and its discussion with other peers in the field for the development and the advancement of novel theories.

Scientific publishing during COVID-19. Communication is done in several ways: classically, we publish our results in peer-reviewed journals, books, or proceedings. Trivially, problems in gathering new empirical findings will result in a slow down of the communication pipeline, which initially might be compensated for by clearing the backlog of data waiting in every researcher's drawers to be published. Paradoxically, however, due to the pressure to publish which most researchers feel to advance their careers, the lockdown during the COVID-19 pandemic might not necessarily lead to a slowing down in the flow of publications, but the limited access to laboratory facilities may well lead to papers of inferior scientific quality [6]. Studies involving specialized equipment, such as VR/AR headsets, might be difficult to distribute to large numbers of people. However, short-term sessions with many users might be replaced with more longitudinal studies with fewer users. This pressure to produce results not only concerns scientific publications but may equally affect thesis projects of bachelor or master students who need to conclude their studies. As discussed earlier in this paper, there are many reasons for which data collection during the lockdown of the COVID-19 pandemic may be harmed, compromised, or even stopped. Regardless, researchers may well be tempted to try and publish this data, knowing (and not concealing) that its quality might be compromised.

As a journal editor, one of our authors has witnessed reviewers that were tempted to adapt their scientific criteria when they learned about the difficulties with acquiring compromised empirical data. Also, the VIS and other communities have offered guidelines asking reviewers to be more forgiving and open-minded of user testing contributions. On the one hand, ethically this is a very laudable move. On the other hand, shifting such criteria might lead to a range of publications which, under other circumstances, might not get published, and which obviously could represent a problem for the scientific process. How big a problem this might be can only be evaluated in the future. The hope is that the processes for ensuring quality, which have been established at journals and in other dissemination channels and which have been critically discussed by responsible editors, remain robust enough to withstand such understandable leniency.

Meetings and conferences. The fact that most publications nowadays are not written by a single author but are co-authored by several collaborators, has likely not affected the publication process significantly, as most of the communication between authors was virtual even well before the COVID-19 pandemic. However, another crucial aspect of disseminating scientific findings has undergone a dramatic metamorphosis in the recent days, namely the communication of research at conferences. Due to the lockdown and the restrictions in traveling world-wide, many conferences have had to be canceled, significantly postponed, or they changed into virtual meetings. Even though many conference organizers have attempted to enable forms of interaction in a virtual format, the scientific experience and particularly the networking between researchers has been severely diminished. The fear is that these restrictions in networking opportunities will particularly impact young and emerging scientist who are currently still building up their network.

On a positive note, the COVID-19 pandemic will clearly help in adapting the focus of dissemination through the use of new digital tools, as well as the availability of more virtual colloquia and online talks. This strengthening of the open science movement will not only benefit scientists but also the general public.

5 Opportunities

As a response to the crisis brought about by the COVID-19 pandemic, the research communities had to restructure their research pipeline, in particular to overcome limitations imposed by the necessity of social distancing. Here, we claim that, despite this disruption of the research environment, scientists can actually leverage these seemingly restricted contexts in order to explore and develop new scientific opportunities.

Overcoming shortcomings of traditional pipelines. First, we suspect that being forced to adopt new methodologies may permanently alter the ways in which research will be carried out. For example, studies might have been previously administered in person simply because some laboratories had more experience with conducting studies with a given protocol. Resistance to changing procedures is especially high when effect sizes are known to be small. Now that some experiments are forced to be moved online, researchers might find that access to larger and more representative samples online can be achieved without a necessary increase in effort. More importantly, they might also find that doing so does not necessarily interfere with the effect of interest. Some unviable research practices may be abandoned altogether, such as filling out pen-and-paper questionnaires while an experimenter is present or relying on convenience samples composed only of lab members. Some researchers might also innovate by embedding research within consumer technology. For example, Saffo et al. [41] showed that HCI-research can be conducted within established gaming applications. We are certain that common research practices have been innovated and economized, affecting the way we conduct user-studies in the long term.

Increasing quality through formalization. Second, the quality of research could increase as a result of the formalization required for large-scale and online deployment of research. For example, online-studies can only be implemented if implicit laboratory procedures, such as the verbal instruction of participants, are explicitly formalized. Studies involving synchronous interactions between experimenter and participants must now also be redesigned to allow for possible replications without the ad-hoc expert input of the experimenter. Such formalization, when possible, would allow for more scalable and quantitative implementations of empirical research, improving the project in terms of objectivity.

Increasing statistical power. Third, findings may be more robust in the future. Online experiments motivates a move from small convenience samples to larger samples [9]. This increase is likely to improve robustness of the statistical procedures because of more appropriate statistical power. We also believe that the process of carefully redesigning studies, while also weighing resources, statistical power, and ethical considerations, has permanently broadened and improved research methods [12].

Increasing generalizability. Fourth, results may be more generalizable due to increased diversity in test setting, presentation media, and population [24]. Remote asynchronous testing could provide more varied test settings, for instance, the user could have more workload due to a child crying in the background, and render robust results that may generalize to more contexts. Users may have to interact with experiments using diverse consumer devices. Presentation parameters, such as stimulus onset timing, must then be treated as a stochastic rather than a deterministic variable.

A shift to online testing also serves the long-lasting aim of diversifying research populations. Moving from a convenience sample to a randomized national-level sample produces more inclusive and diverse samples, e.g. by including the elderly, non-academics, or physically disabled people who could not visit a laboratory, which allows for a better generalizability across populations [34]. While such scalability offers exciting new opportunities, it must be noted that its full potential can only be realized through careful attention to issues concerning accessibility for participants. Researchers must take into consideration the respondents' capacity to use technology autonomously and address potential impairments or disabilities that would hinder full participation in the study.

A large portion of existing empirical visualization research has focused on the goal to generalize across user characteristics, i.e., a method should work similarly well for many different users. With a constraint access to users for testing, however, researchers might also increasingly think about opportunities to generalize findings across other factors such as data characteristics. Testing visualization methods on larger sets of data can help to make them more robust for broader usage of a method [46]. In other words, a method should work similarly well for many different datasets as well.

Of course, it is important to keep in mind that generalizability is not the only goal of visualization evaluation. In fact, there has been a long tradition in VIS and at BELIV to underline the value of qualitative research methods for visualization research [10, 38]. Those qualitative methods often seek to authentically understand specific and contextualized situations in which visualizations are used and, thus, generalizability, control, and reproducibility are not the concern of these studies. Approaches such as semi-structured interviews should carry over more easily to remote setups through teleconferencing and other techniques. This gives an opportunity to the VIS community to recall the strengths and importance of these approaches.

Strengthening the community. Conversely to social distance requirements, which have halted physical scientific conferences and travel for scientific exchange, it may also be argued that research communities have made a significant step to moving closer together in the crisis. Online conferences have allowed for relatively unrestrained interaction among scientists. Political, economic, and physical barriers can also be avoided through the use of online conferences, which have often been opened to free participation. We suspect that the popularity of online conferences will last far beyond the time of crisis [1]. Due to the problem of recruiting participants for user studies, the crisis has also added pressure to form alliances among researchers to share research data and expertise, as well as test at multiple sites, in order to create more innovative research designs. Borders in scientific exchange and collaboration have thus been torn down. In a similar vein, the workplace of scientists has moved from a local to a remote setting. Scientific careers are typically built upon moving between different scientific groups and thus different places of work. While permanent remote-working is often accepted in online business or web development, academia has mandated local collaboration thus far. We believe that working models such as a remote postdoctoral phase will become more common now that better acceptance and better infrastructure has made long-term remote working possible [8].

This example already highlights that people may be impacted in different ways with regards to their career levels. The career progress of senior researchers can be largely unaffected by social distance requirements and the associated, as unfinished projects could be written-up, questions could be rephrased, and contracts extended. However, early career researchers may have not yet collected data, or they are in need of local support and may require access to facilities in order to run their studies. The community should therefore keep in mind that early career researchers are more vulnerable to the effects of the COVID-19 crisis.

The situation concerning education and the necessary restructuring of university programs deserves further consideration. Due to the shifting of many of the university activities online, many younger researchers and students, in particular in the early stage of their careers, are reconsidering their timelines and possibly delaying their studies. This shift in the demographics of incoming researchers in virtually all scientific fields will most likely lead to larger structural changes in the research community itself.

Opportunities for the VIS community. Concerning the visualization pipeline specifically, moving studies toward crowd-sourced setups can be an opportunity with wide repercussions. While lab studies offer controlled setups and ensure that users complete the experiment with meaningful answers, the participant availability often leads to new visualization techniques only being tested on few toy datasets. If the findings generalize well across user bases, it is questionable whether they generalize well across datasets [19, 46].

We also see an opportunity to develop active learning or labeling pipelines developed in the machine learning community [47]. The visualization community traditionally focused on improving the labeling process for humans [4, 5, 22, 28]. Perhaps it is time to take our own medicine and aid visualization evaluation by having algorithms learn which stimuli have to be evaluated by humans and which can be evaluated by machines.

A crowd-sourced setup aligned with active labeling systems can enable testing datasets found in the wild and can help evaluating designs on more realistic datasets, and with more diverse users, ultimately leading to more robust results [18].

6 Conclusions

COVID-19 had a huge effect on user-centered testing in particular and entire research pipelines more generally. In this work, we tried to take a first step to understanding challenges, potential solutions, and even opportunities from a visualization point of view. The pandemic might have far-reaching implications and, as a community, might encourage us to re-think and change our existing norms and value systems. Of course, these questions are far beyond a single paper. We see our work as a starting point for such discussions and hope that others will find some motivation and inspiration in the topics we raised.

Acknowledgments

We thank Christopher Geekie for the useful comments on the language and style of the article. This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161.

References

[1] T. Achakulvisut, T. Ruangrong, I. Bilgin, S. Van Den Bossche, B. Wyble, D. F. Goodman, and K. P. Kording. Improving on legacy conferences by moving online. *eLife*, 9:1–4, 2020. doi: 10.7554/ELIFE.57892

[2] J. P. Andersen, M. W. Nielsen, N. L. Simone, R. E. Lewiss, and R. Jagsi. Meta-Research: COVID-19 medical papers have fewer women first authors than expected. *eLife*, 9:1–7, 2020. doi: 10.7554/eLife.58807

[3] R. Arias-Hernandez, L. T. Kaastra, T. M. Green, and B. Fisher. Pair analytics: Capturing reasoning processes in collaborative visual analytics. In *2011 44th Hawaii International Conference on System Sciences*, pp. 1–10. IEEE, 2011.

[4] M. Behrisch, F. Korkmaz, L. Shao, and T. Schreck. Feedback-driven interactive exploration of large multidimensional data supported by visual classifier. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 43–52. IEEE, 2014.

[5] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):298–308, 2017.

[6] L. Besançon, N. Peiffer-Smadja, C. Segalas, H. Jiang, P. Masuzzo, C. A. Smout, M. Deforet, and C. Leyrat. Open science saves lives: Lessons from the COVID-19 pandemic. *bioRxiv*, 2020. doi: 10.1101/2020.08.13.249847

[7] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of eye tracking data: A taxonomy and survey. *Computer Graphics Forum*, 36(8):260–284, 2017. doi: 10.1111/cgf.13079

[8] K. R. Burgio, C. M. MacKenzie, S. B. Borrelle, S. K. M. Ernest, J. L. Gill, K. E. Ingeman, A. Teffer, and E. P. White. Ten simple rules for a successful remote postdoc. *PLOS Computational Biology*, 16(5):1–9, 2020. doi: 10.1371/journal.pcbi.1007809

[9] K. Caine. Local standards for sample size at chi. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, p. 981–992, 2016.

[10] S. Carpendale. Evaluating information visualizations. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, eds., *Information visualization*, pp. 19–45. Springer, 2008.

[11] A. Crisan and M. Elliott. How to evaluate an evaluation study? comparing and contrasting practices in vis with those of other disciplines: Position paper. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pp. 28–36. IEEE, 2018.

[12] P. Dragicevic. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*, pp. 291–330. Springer, 2016. doi: 10.1007/978-3-319-26633-6_13

[13] M. B. Eisen, A. Akhmanova, T. E. Behrens, and D. Weigel. Publishing in the time of COVID-19. *eLife*, 9:9–11, 2020. doi: 10.7554/elife.57162

[14] D. Feil-Seifer, K. S. Haring, S. Rossi, A. R. Wagner, and T. Williams. Where to next? The impact of COVID-19 on human-robot interaction research. *ACM Transactions on Human-Robot Interaction*, 10(1):1–7, 2020.

[15] G. A. Gescheider. *Psychophysics: The Fundamentals*. Psychology Press, 2013.

[16] S. R. Gomez, R. Jianu, R. Cabeen, H. Guo, and D. H. Laidlaw. Fauxvea: Crowdsourcing gaze location estimates for visualization analysis tasks. *IEEE Transactions on Visualization and Computer Graphics*, 23(2):1042–1055, 2016.

[17] S. Haroz. Open practices in visualization research: Opinion paper. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pp. 46–52. IEEE, 2018.

[18] J. Heer and M. Bostock. Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 203–212, 2010.

[19] K. Hu, S. Gaikwad, M. Hulsebos, M. A. Bakker, E. Zgraggen, C. Hidalgo, T. Kraska, G. Li, A. Satyanarayan, and Ç. Demiralp. Viznet: Towards a large-scale visualization learning and benchmarking repository. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[20] P. Isenberg and S. Carpendale. Interactive tree comparison for colocated collaborative information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1232–1239, 2007.

[21] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics*, 23, 2017.

[22] E. J. Keogh and M. J. Pazzani. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, vol. 98, pp. 239–243, 1998.

[23] N. W. Kim, Z. Bylinskii, M. A. Borkin, K. Z. Gajos, A. Oliva, F. Durand, and H. Pfister. Bubbleview: an interface for crowdsourcing image importance maps and tracking visual attention. *ACM Transactions on Computer-Human Interaction*, 24(5):1–40, 2017.

[24] R. A. Klein, M. Vianello, F. Hasselman, B. G. Adams, J. Reginald B. Adams, S. Alper, M. Aveyard, et al. Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4):443–490, 2018.

[25] R. Kosara and S. Haroz. Skipping the replication crisis in visualization: Threats to study validity and how to address them: Position paper. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pp. 102–107. IEEE, 2018.

[26] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2011.

[27] J. Lazar, J. H. Feng, and H. Hochheiser. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann, 2017.

[28] F. Lekschas, B. Peterson, D. Haehn, E. Ma, N. Gehlenborg, and H. Pfister. Peax: Interactive visual pattern search in sequential data using unsupervised deep representation learning. *Computer Graphics Forum*, 39(3):167–179, 2020.

[29] Q. Li, S. J. Joo, J. D. Yeatman, and K. Reinecke. Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific Reports*, 10(1):1–11, 2020.

[30] L. Litman and J. Robinson. *Conducting Online Research on Amazon Mechanical Turk and Beyond*, vol. 1. SAGE Publications Inc., 2020.

[31] S. F. Lourenco and A. Tasimi. No participant left behind: Conducting science during covid-19. *Trends in Cognitive Sciences*, 24(8):583 – 584, 2020.

[32] J. E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In R. M. Baecker, J. Grudin, W. A. Buxton, and S. Greenberg, eds., *Readings in Human–Computer Interaction*, pp. 152–169. Elsevier, 1995.

[33] A. J. Moss, C. Rosenzweig, J. Robinson, and L. Litman. Demographic stability on Mechanical Turk despite COVID-19. *Trends in Cognitive Sciences*, 24(9):678–680, 2020.

[34] A. Mottelson and K. Hornbæk. Virtual reality studies outside the laboratory. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, 2017. doi: 10.1145/3139131.3139141

[35] T. Munzner. Process and pitfalls in writing information visualization research papers. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, eds., *Information visualization*, pp. 134–153. Springer, 2008.

[36] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.

[37] K. Myers, W. Y. Tham, Y. Yin, N. Cohodes, J. G. Thursby, M. Thursby, P. Schiffer, J. Walsh, K. R. Lakhani, and D. Wang. Quantifying the immediate effects of the COVID-19 pandemic on scientists. *SSRN Electronic Journal*, 2020. doi: 10.2139/ssrn.3608302

[38] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced Visual Interfaces (AVI)*, pp. 109–116, 2004.

[39] S. Porter and D. Hook. How COVID-19 is changing research culture, 2020. doi: 10.6084/m9.figshare.12383267.v2

[40] C. W. Rudolph, B. Allan, M. Clark, G. Hertel, A. Hirschi, F. Kunze, K. Shockley, M. Shoss, S. Sonnentag, and H. Zacher. Pandemics: Implications for research and practice in industrial and organizational psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2020.

[41] D. Saffo, C. Yildirim, S. Di Bartolomeo, and C. Dunne. Crowdsourcing virtual reality experiments using vrchat. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, p. 1–8, 2020. doi: 10.1145/3334480.3382829

[42] M. Sauter, D. Draschkow, and W. Mack. Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, 10(4):251, 2020.

[43] A. Schmidt and F. Alt. Evaluation in Human-Computer Interaction-Beyond Lab Studies.

[44] N. D. Schwartz, B. Casselman, and E. Koeze. How bad is unemployment? 'Literally off the charts'. *The New York Times*, 8, 2020.

[45] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.

[46] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A taxonomy of visual cluster separation factors. *Computer Graphics Forum*, 31(3pt4):1335–1344, 2012.

[47] B. Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.

[48] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–7, 2006.

[49] A. M. Simmons and S. J. Luck. Protocol for reducing COVID-19 transmission risk in EEG research. *Research Square*, 2020. doi: 10.21203/rs.3.pex-974/v2

[50] D. Toker, C. Conati, B. Steichen, and G. Carenini. Individual user characteristics and information visualization: connecting the dots through eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 295–304, 2013.

[51] N. S. Wigginton, R. M. Cunningham, R. H. Katz, M. E. Lidstrom, K. A. Moler, D. Wirtz, and M. T. Zuber. Moving academic research forward during COVID-19. *Science*, 368(6496):1190–1192, 2020. doi: 10.1126/science.abc5599

[52] J. T. Wixted and S. L. Thompson-Schill. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Language and Thought*, vol. 3. John Wiley & Sons, 2018.

[53] World Health Organization. Ethical standards for research during public health emergencies: distilling existing guidance to support COVID-19. *Ethical standards for research during public health emergencies: Distilling existing guidance to support COVID-19 R&D*, pp. 1–4, 2020.