Perspectives on Negative Research Results in Pervasive Computing

Ella Peltonen [ID], *University of Oulu, 90014, Oulu, Finland*

Nitinder Mohan [ID], *Technical University Munich, 00014, Munchen, Germany*

Peter Zdankin, *University of Duisburg-Essen, D-45117, Duisburg, Germany*

Tanya Shreedhar [ID], *University of Edinburgh, EH8 9YL, Edinburgh, U.K.*

Tri Nguyen [ID], *University of Oulu, 90014, Oulu, Finland*

Suzan Bayhan [ID], *University of Twente, 7522 NB, Enschede, The Netherlands*

Jon Crowcroft [ID], *University of Cambridge, CB2 1TN, Cambridge, U.K.*

Jussi Kangasharju [ID], *University of Helsinki, 00100, Helsinki, Finland*

Daniela Nicklas, *University of Bamberg, 96047, Bamberg, Germany*

Not all research leads to fruitful results; trying new ways or methods may surpass state of the art, but sometimes the hypothesis is not proven, the improvement is insignificant, or the system fails because of a design error done years ago in previous works. In a systems discipline like pervasive computing, there are many sources of errors, from hardware issues over communication channels to heterogeneous software environments. However, failure to succeed is not a failure to progress. It is essential to create platforms for sharing insights, experiences, and lessons learned when conducting research in pervasive computing so that the same mistakes are not repeated. And sometimes, a problem is a symptom of discovering new research challenges. Based on the collective input of the First International Workshop on Negative Results in Pervasive Computing (PerFail 2022), co-located with the 20th International Conference on Pervasive Computing and Communications (PerCom 2022), this article presents a comprehensive discussion on perspectives on publishing negative results, useful failures, and lessons learned in pervasive computing.

Research in fields such as pervasive computing, networking, systems, etc., is usually driven by results that push forward (or improve the efficiency of) state-of-the-art or provide demonstrable contributions. To achieve novel and practical solutions, researchers often invest significant efforts into designing and evaluating several iterations of design choices with the hopes of achieving the desired performance. However, the probability of such efforts leading to a negative or failing output can be quite high due to a lack of complete understanding/hindsight or unconvincing design choices, some of which may be harder to rectify at later stages. Such efforts are rarely rewarded since negative results or lessons learned are usually hard to publish,[1] being considered not novel or lacking significant new knowledge. However, we argue that if properly leveraged, negative results and lessons learned can be beneficial to the research community at large from the perspective of knowledge transfer between the research groups tackling the same topics. While our arguments in this article are from the perspective of pervasive computing research, many of our insights are valid in other fields of computing, such as human–computer interaction, network/mobility measurements,

security and privacy, etc., thanks to the inherently interdisciplinary nature of pervasive computing.

Most pervasive computing systems require multiple, often nonstandard components, like hardware prototypes for sensing, developing computation paradigms, novel energy-saving communication protocols, adaptive middleware, or multimodal user interfaces. Often, a new research idea or hypothesis covers only one part of the system, but to complete the experiment, you must (at least partially) get everything to work. Each component might be a source of error or the bottleneck that prohibits the expected result. However, sometimes the interplay of different components highlights the symptom of an underlying, more significant challenge that could lead to new research ideas. Hence, in this article, we assert that we *must* treat negative results and other failing experiments as a potential source for new scientific insights.

While unsuccessful outcomes might not directly contribute to the advancement of the state-of-the-art, the wisdom of hindsight could be an essential contribution in itself, such that other researchers could avoid falling into similar pitfalls. Such "what not to do?" insights can also foster good research practices, especially for handling gray areas like ethical boundaries.[2] We consider negative results to be outcomes of studies that are run correctly (in the light of the current state-of-the-art) and in good practice but fail in proving the hypothesis (statistically significantly). In addition, essential lessons learned can also emerge from properly but unfittingly designed data collection or (nontrivial) lapses of hindsight, especially in studies involving real-world uncontrolled measurements. Furthermore, the experiences that help spot adverse outcomes from intermediate results benefit the pervasive computing community at large.

Besides, it is crucial to share (especially for junior researchers) the knowledge that not all experiments will be successful, failures might happen, and knowing how to overcome unexpected outcomes is a skill to develop as a researcher. The interest in this field was evident from the participation in the questionnaire for Percom 2022 participants, in which $\approx$70 attendees participated. The result of the questionnaire revealed that $\approx$85% of the conference participants had already failed in their research work. The perspectives noted in this article are drawn from discussions with the pervasive computing community during the First International Workshop on Negative Results in Pervasive Computing (PerFail 2022)[a] and from authors' past experiences. Specifically,

this article discusses failure taxonomy in pervasive computing, considers mitigation and avoidance strategies, and presents our call for arms for healthy failure culture in pervasive computing research.

WHAT IS A FAILURE?

Traditional applied research, including pervasive computing, aims to fundamentally understand how systems work or how to make them work. Within the pervasive computing, systems and networking research, solutions that *demonstrably* work are naturally more attractive to the community. Therefore, it becomes essential to gauge the usefulness of research that failed to work. Failure can be viewed as a necessary catalyst to research—if there is no possibility of failure, the novelty of the research is also questionable.

Within the research ecosystem, the definition of "failure" is often misconstrued, which makes many researchers incorrectly label unexpected outcomes of research studies as failures. Since the underbelly of the ecosystem is somewhat driven by "publish or perish" ideology,[3] researchers tend to inherently add significant weight to the number (and quality of venue) of publications. However, we argue (on logical grounds) that publications are means for disseminating new-found results; hence failing to publish in a competitive venue or securing funding should not be considered as failed research. While the scientific community inadvertently uses specific criteria to measure a researcher's success—citations, publications at high-impact venues, funding acquired–these may not fully reflect the success of the conducted research. Instead, the criteria reflect how the output is *perceived* by the scientific community in the short term. For example, a publication about a specific failure may be successful on its own by preventing others from repeating said failure, hence having a significant impact on the research that is performed. But this publication has a low chance of being cited in the hopefully successful research that it leads toward. As such, we first discuss different categories of results obtainable from research studies.

Categories of Research Results

I. Positive results are the results that show the validity of the research hypothesis. Studies with positive results are more likely to be cited, indicating a bias in the community in publishing works showcasing positive results.[4,5]

II. Negative results are the results that contradict previously established knowledge (e.g., lack of statistical significance to prove the hypothesis of these earlier

---

[a][Online]. Available: https://perfail-workshop.github.io/2022/

studies or statistically significant results contradicting the earlier positive results). Publishing negative results can clarify and contradict controversial claims,[b] and bring deeper understanding into novel results that sounded too good to be true.[c] In pervasive computing, we rarely have controversial claims that make headlines. However, before disclosing contradicting (either positive or negative) results, it is crucial to have confidence that the research is rigorously designed and executed.[6] Careful consideration of research methods and study design (as discussed later in this paper) plays a significant role in avoiding the waste of resources (e.g., researchers' time or public money).

In addition, negative results can prove a unique collaboration opportunity for involved research groups to understand the possible reasons for the discrepancy between the obtained results. Negative results are also related to reproducibility. Recreation of previous studies might be affected by uncontrollable configuration changes, design, and assumptions, which can lead to different conclusions. Therefore, it is paramount to follow open science practices to facilitate a more accessible investigation of the root causes of discrepancy.[7] According to a small-scale survey with 96 researchers found that the majority of the respondents saw value in sharing the adverse research outcomes, but did not attempt publishing due to considerable time and effort investment with potentially less number of citations compared to the studies with positive results.[8]

III. Unexpected or unexplainable results are results that cannot be explained with the hypothesis being investigated, requiring further exploration and reassessment of problem formulation, assumptions, or design choices. Unexpected results are likely caused by flaws in methodology (e.g., software errors) or unrealistic assumptions but might also emerge due to the complex system interactions not captured by the hypothesis or designed experiments. For example, in de Vos et al.[9] the authors highlight that the reported behavior of distributed system solutions in a controlled testbed can be significantly different from a large-scale deployment—with the latter uncovering several performance artifacts that authors may remain oblivious to otherwise. In pervasive computing, many factors from the used sensors, their calibration, and the composition of study participants, play a significant role and can lead to a mismatch between expected and obtained results. It is, however, crucial to understand

that bad research–e.g., wrongly done study design, hypotheses that do not make sense, and bad work caused by incompetence–is not a learnable failure we seek in this article. This is why we discuss the cycle of the research and failure taxonomy later in this article.

Not all research studies contain results that can be proved or disproved. *Vision papers* raise important questions pointing to new possibilities and draw an agenda pioneering a field or vision for researchers to pursue. For opening new avenues and out-of-box thinking beyond the state-of-the-art, vision papers play a crucial role in advancing science.[10] Similarly, *survey studies* and *tutorial papers* can provide a good understanding of the field and assist researchers in identifying the knowledge gaps.

The Cycle of Research

We now discuss the cycle of research that can potentially lead to such different achievable results, illustrated in Figure 1. To conduct research that advances the field and is impactful, researchers should raise the "right" questions to work on. Finding good research questions that fill a knowledge gap may be challenging, as research directions that lie at the intersection of being solvable and pertinent to the community can be limited or require a deep understanding of the literature. At this stage of *formulating the research questions*, apart from performing an extensive literature review with an inquisitive mind, researchers can benefit from discussing hypotheses with colleagues (in addition to advisors) or in discussion forums such as Ph.D. schools, workshops, and community meetings.

Note that performing actions within this stage in conjunction with the next stage of the cycle, i.e., plans to execute the research, is likely to yield desired positive outcomes. For example, suppose the research question involves understanding user behavior over certain stimuli. In that case, it is beneficial to prototype design/methodology for conducting the study, enumerate parameters that need to be recorded and the expected results from the campaign. It is pertinent that researchers are not rigid with their initial assessment of the suitability of the research questions and means to achieve it to avoid *cognitive bias*—that can significantly impact the outcome of the research. As such, they should be willing to consider preliminary feedback and *reassess* their approach, formulation, and (if required) research direction—since realigning at these initial stages can maximize research impact with the least negative consequences.

Once the research formulation is finalized, researchers should periodically assess the intermediary results from iterative prototypes for revisions, if

---

[b][Online]. Available: https://www.nature.com/articles/nature.2012.10249
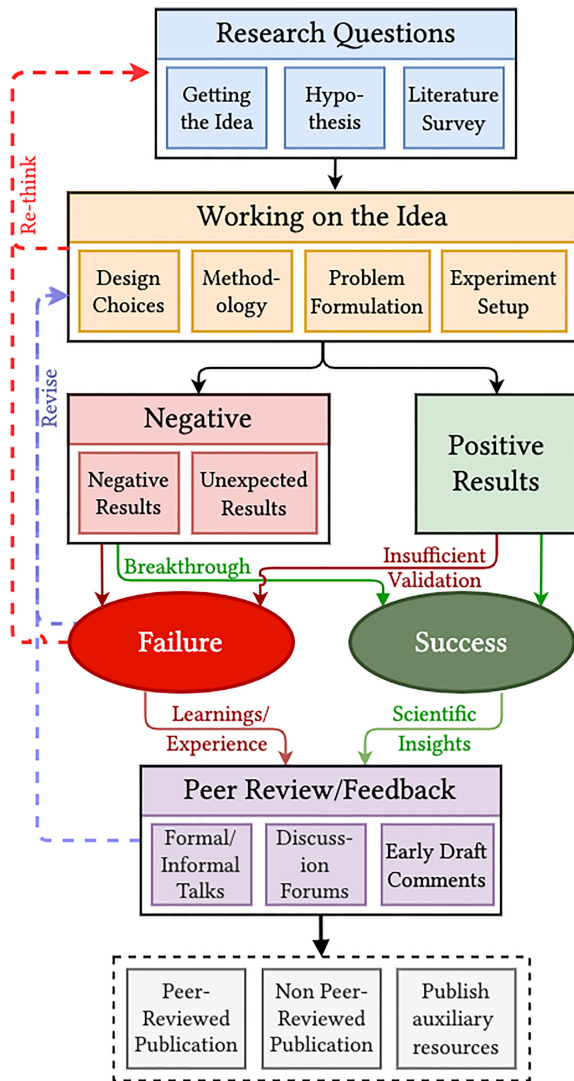[c][Online]. Available: https://www.nature.com/articles/d41586-019-01675-9

FIGURE 1. Cycle of research.

necessary. Most importantly, it is essential to remember that the experiments can likely yield negative or unexpected results at the end of the study. At this stage, researchers must discover the reason behind their unexpected results—is it due to wrongful formulation or lack of collected parameters, or is it challenging the widespread assumption within the field? As such, a negative outcome can result in a successful study with a *breakthrough*, albeit a strong validation proves no shortcomings/bias in methodology and problem formulation. Similarly, a positive result from a study does not immediately equate to a successful outcome and requires statistically significant validation to prove no bias in the experiment formulation.

In the next step of the cycle, the researchers usually formulate their findings into research articles and

submit them to peer-review venues. It is important that at this stage, enough checks are performed by the researchers, such that they are confident that their results are of positive or negative nature. A high-quality peer review will ease determining the number of attempts required. While a significant portion of the scientific community questions the accuracy of the current peer-review systems, often terming it to be "broken,"[3] the core utility of the process is to ensure that the submitted research is free from overlooked biases/errors and incorrect information.

It must be noted that despite widely accepted notions, publications are not the *goal* of a research study but merely means for disseminating the new-found knowledge or experiences gained during the course of the research. Successful research can also produce practical systems and artefacts, which do not require traditional peer-review conference/journal publications but open-source code with descriptive documentation, website, or even blog post. As a result, researchers should aim to engage with the community in alternate methods (in addition to research publications) to receive a holistic peer review and feedback on improving the research. For example, networking and systems research outcomes can include RFC drafts to be discussed in venues like IETF/IRTF (which may also result in failures[11]). Researchers can give formal or informal talks at conferences and research institutes and advertise early results in mailing lists, community forums, Slack channels, and so on.

Expert peer review of any form (traditional publication as well as via alternate means) must be considered with due diligence, which might uncover earlier overlooked sources of errors in the research methodology. As a result, the researchers should be open to *revise* their research methodology and formulation to address correctness concerns.

Despite repeated improvements, it might be possible for the manuscript to be repeatedly rejected from multiple publication venues for various reasons. The phenomenon is often misrepresented as a "failed research." However, as previously stated, failing to publish a paper at the desired venue does not necessarily point to insufficient quality of the research study. As such, researchers should avoid the mindset of giving up prematurely after experiencing a string of publication rejections. Different publication venues have different approaches and understanding of the methodology and research approaches; hence, choosing the venue correctly and ensuring that the manuscript is well-written can be crucial.

Researchers should also consider submitting to less competitive venues to maximize the *impact* and

timeliness of a research study. For example, a study that facilitates further research with more efficiency (e.g., via developing a measurement tool) can be as impactful as studies with groundbreaking ideas. Moreover, the impact of many studies can become evident over time with the advances in supporting sciences and technology. It might turn out that peer-review feedback or methodology revisions reveal that research has failed or obtained insignificant results. Sharing lessons learned from such failed experiments will also help advance the field's understanding, as others can avoid similar pitfalls that impact the study. However, care must be taken to accurately summarize the sources of errors/learning and potential future directions for improvement, which must not be mistaken for publishing incorrect results.

The researchers can also look for venues that actively publish and discusses failure results catering to their research field, e.g., PerFail, Insights, and JOURNE—and more are welcomed in the field. Knowing the possible failure points in the research cycle, knowing the right community for the work in question, and writing an article that is acceptable to the relevant community only come by experiencing research. The question, in essence, is not how to avoid failure—more important is to understand what the impacts of the performed research are, negative or positive, and disseminate them. Learning from failures is one of the key skills supervisors can deliver to their juniors and research communities for their members.

FAILURE TAXONOMY

It is important to understand what are the key lessons learned and failures worth sharing with the research community. Thus, in this section, we identify what a failure is and its key characteristics that hint at research outcomes that might fail. The purpose of this section is not to guide the reader not to fail; on the contrary, we underline that every researcher will fail at some point in their career, and many do it multiple times. The first step in communicating these failures is, however, recognizing them. In addition to naturally avoiding failures, it could be a successful idea to consider the mitigation of failures that are not always evident. The failure mitigation measures include changing the research direction, study design, or methodology early on, gathering insights, and publishing the negative results and lessons learned if the research fails at a later stage.

The Origin of Failures

Research should start with a clear problem statement and a verifiable hypothesis, as a lack of either is a well-

known predetermined breaking point for the subsequent research. One of the most critical steps in producing "*good*" research is a comprehensive understanding of the problem that is to be solved. If the problem is not understood well enough, it is possible to research something that lies close to a problem and perhaps is an interesting field on its own but fails to produce fruitful results that can be published. While open-ended research in a preliminary phase is useful for collecting ideas, it eventually should be fitted into a more solid framework with clearly-defined goals. Experience says that the sooner the failure is identified, the easier it is to fix; this is true in all empirical sciences, including pervasive computing.

Related work is critical to advance in science, as this step encourages findings on this topic made by other researchers. This could include positive or negative results on a similar topic, to either be a warning of spending time and effort on a problem that is either unsolvable or already solved. Even more important, the time frame of scanning the related work should not be too small, and perhaps 10, 20, or more years ago, a problem was already solved. This is especially true with somewhat new research fields such as pervasive computing; on the other hand, the field focuses on rapidly advancing technology, where there is a notable amount of new research to cover. On the other hand, many application areas existed before the term "pervasive computing," and solutions may have been found earlier than the field's emergence.

New Research Topics

It is easy to focus on the wrong part of the problem or misunderstand the significance of the results, either because you are looking at incomplete or incorrect information or trying to interpret different things from the results from what they actually show. Being early on a topic, especially a topic that later becomes *hot* such as deep learning or blockchains lately, is what all researchers dream of doing. But being early on a new topic means there is not an established community around, so there might be no general understanding of the significant results or the critical questions in this area. This might seem counter-intuitive because research is about discovering new things, and somebody always has to write the first paper on a new topic. There is an excellent scope of vision papers in such scenarios.

Failure Types

Failures come in all kinds of shapes and forms (see Table 1). Following previous overviews of the experiences and lessons learned while conducting research,[12,13]

TABLE 1. Summary of potential failure points at different stages of a research study.

	Failure Points	Recommendations
Problem Formulation	Wrong hypothesis Being ahead of time Unclear research questions Low relevance to venue Lack of novelty or delta increment to existing works	Clear demarcation from literature Vision papers Extensive literature review
Study Design and Execution	Theory vs practice Unrealistic assumptions Too simple or too complex abstraction	A clearly-described methodology section Citing relevant resources Detailed model descriptions
Measurements	Software/Hardware errors Human/sensors errors	Avoiding ambiguity Open datasets Community accepted methodology
Performance Analysis	Wrong metrics and methodology	Comparison with state-of-the-art
Publication	Repeated rejections	Re-assess relevance and publication venue

we identify the following set of failure types relevant to the pervasive computing field:

› Unconvincing results which could not be verified because it is not even possible, for example, due to a lack of datasets, real-life experiments, or ground truth.

› Under or overperforming experiments that could have been possibly fixed, for example, oversights in system design, inadequate or misconfigured infrastructure, or just buggy code.

› Setbacks resulting in lessons learned and acquired hindsight, including hypotheses with too limiting or too broad assumptions.

› Unconventional, abnormal, or controversial results that contradict community expectations and are thus extremely hard to prove reliably.

› "Non-publishable" or "hard-to-publish" side outcomes of the otherwise correctly run studies, including mistrials of experiment methodology or design, preparations for proof-of-correctness of results, and so on. The limited contribution might not be a failure itself, but might become hard to publish as a full research article.

› Unexpected roadblocks affecting publications, including ethical concerns, institutional policy breaches, and difficulties in collaboration with commercially bound parties of the research projects. Ethical and privacy hurdles might only be encountered after a series of measurements have already been conducted, forcing an unwanted halt to the entire research, including any observations made prior to any private data being collected.

FAILURE MITIGATION

In previous sections, we have discussed where, when, and how failures in pervasive computing research may occur. More importantly, we want to highlight that there is no researcher free of failure. One of the key ideas behind our workshop has been to systematically understand these failures and share lessons learned from them, preventing others from doing the same mistakes. During the PerFail workshop, we covered failures that have happened in simulations, performance analysis, and the more technical side of the field. Some of the key lessons learned discussed in the workshop include:

› *Use step-wise study procedures*, failures and setbacks are easier to recognize during the study.[14,15]

› *Design for the target group and consult domain experts*, especially if the study concerns human subjects. Seek feedback from domain experts for any modification to the questionnaire.[14]

› *Prioritize collecting high-quality data* in the firsthand. Consider the quality of labeling the data items and events, especially when collecting the data in the wild. The sensor devices have to ensure the quality of the signal, even though existing noise.[14,15]

› *Ensure validity of the ground truth data* by considering "gold standard" devices, especially with sound studies evaluating self-gathered field data with smartphones and wearable devices. The ground truth data must be validated in the lab and in the wild environments. Data collection can be monitored via a dashboard to filter noise and abnormality.[15]

› *Avoid ambiguity:* If the study includes real people, accommodate your research for the situation. If using questionnaires, create them in a self-explainable way to avoid misunderstandings.[15] In their IETF experience postmortem,[11] reveal that ambiguity in describing intended protocol behavior in different circumstances can also cause failures in standardization bodies.

> *Relation between data selection and experimental plans:* Based on the requirements and practical techniques, selecting a fit dataset affects the final results of the research.[16] Moreover, the deployment size of the experiment can also affect outcome.[9]
> *Analyze previous methodologies for deep comparison and understanding:* After deciding on a fit dataset, understanding the dataset needs an effort in analyses from earlier works for data processing and methodologies. Understanding the dataset for experiments is to avoid a lack of important features that fit the experimental requirements.[16]
> *Consider resources related to the experiments:* The experimental results mainly indicate the boost in a feature; however, the tradeoff from other aspects or resources needs to be carefully considered, especially when transforming methodologies into a new environment.[17]

In addition, careful consideration of ethical guidelines is essential.[18] This concern may not only protect the integrity of the subjects of a study but may also prevent the halt of entire studies. To better safeguard any research involving personal data, the use of Institutional Review Boards is strongly advised and even often required by outlets.[19] Identifying the failures after the manuscript has passed the review cycle and got published is not only more than morally right but also crucial for the reliability of the research. Publishing a retraction, errata, or follow-up are all critical tools to avoid just "living with" an error that could have been fixed. This has to be the case whether the error is recognized by the original authors or someone coming after them (who might, surely, publish a new paper proving the previous results wrong).

Exciting and risky research will necessarily have more failures than successes; this needs to be understood and normalized in scientific publications. Unlike industry and startup projects, not every research project must yield a success story, as even the insight of a methodology not working for a given problem is a revelation on its own.[17] There are some research directions that are both risky and expensive, where a registered experiment publication (e.g., RSOS[d]) or any publication about negative results is vital to prevent such sunk costs from happening again.

---

[d][Online]. Available: https://royalsocietypublishing.org/journal/rsos

In addition, it is notorious that there are unexpected outcomes or a hypothesis that does not hold.[20] In these cases, it is crucial to show the importance of negative results and lessons via parameters, including the design of experiments, evaluation of performance, comparison of models, statistical hypothesis testing, and so on. It needs to be emphasized that there is a difference between negative results and no results, that the no-results indicate incomplete or unreliable works. Meanwhile, the negative results contain valuable information; they should be welcomed to save efforts in redundant works and contribute intellectual side related to models, methods, and datasets.

A CALL-FOR-ARMS FOR HEALTHY FAILURE CULTURE IN RESEARCH

Even the most experienced researchers can come up with a research question and hypothesis that proves wrong. If the outcome is not expected or aimed for, is this a failure? More importantly, how should it be reported in the community that the first hypothesis did not work? Publishing failures and negative results enable others to avoid making the same mistakes, yet there are significantly more nonworking solutions than good, efficient ones.

Most research is what is commonly called "incremental," i.e., it involves small perturbations of previous research. Done correctly, it entails reproducing the earlier results and adding some variation. Even if the variation does not yield the expected results, it has successfully reproduced and thus validated the earlier work. In our opinion, this is definitely not a failure. Unfortunately, too often, such "failed" variations are reported as successes by sugarcoating the results or trying to spin them as positive results. Partly this is because we, as a research community, expect good and positive results. We should be more tolerant and recognize that reproducing earlier results and reporting negative findings also carries significant value.

Some research aims at creating something radically new and different or finding completely novel research directions. In most cases, these end up failing. Some funding agencies (e.g., European Research Council) explicitly call for *high risk, high gain* types of proposals, while others focus (implicitly) on more incremental work with more guaranteed but smaller successes. We should understand that both kinds of research have value and that failures may simply be indicators of the level of ambition. To propagate a healthier failure culture in research, we propose a call-for-arms action in the community (with a broader relevance outside pervasive computing).

"Guardians of Publication Quality"

The review process itself is not free of criticism. What will be accepted is, by human nature, dependent on the opinion of reviewers, accompanied by the game of chance to get the reviewers assigned that view the publication most positively.[3] The same game of chance is present in the unregulated nature of paper reviews, which allows for claims and unfairness by the reviewers if the editors do not secure uniform quality reviews.

The task of a reviewer or editor *should not* be about improving the prestige of the publication venue, but instead checking the correctness of the work—to the extent feasible. Of course, not all problems in the review process are due to "malicious" reviewers. The difficulties are real in conveying the applicability of studies regardless of their limited metrics and comparing publications against each other, as the performance metrics applied to decide the fate of one submission cannot be used for another. In addition, there is a constant need for novel openings instead of building on old theories, despising potential but risky new research fields. Highly selective conferences tend to favor "exciting" results, leaving little to no room for the "correct but boring" work that has been done rigorously yet falls under the not-so-successful variation of earlier work. As reviewers determine the fate of a paper, without a change in the reviewers' mindset, it is hard to see papers with negative results getting their spot in publication venues. Hence, a common understanding and guidelines should be developed for more fair treatment of such studies similar to recommendations provided in the San Francisco Declaration on Research Assessment (DORA).

Yet, an interesting question is how much publication space should failures or negative results get? If the result refutes earlier work, it deserves its spot in the limelight. However, if it simply highlights that a particular variation/configuration did not work, how should it be communicated? Should publication venues dedicate some space to "failures," dictating authors to report such unexpected results in an appendix or a specifically reserved section? Or should we have dedicated publication venues for publishing such negative results? The former allows readers to learn the underbelly evolution of the work alongside positive results, providing a new learning perspective to the research study. On the other hand, the latter would give a general prominence to "things that did not work" in various open problems in the field, encouraging people to take more risks, knowing they have a good chance of advancing the state-of-the-art through unexplored approaches.

Normalize Failures as Researchers

Our aim, through this article, is to normalize failures as part of the research process and motivate their means of communication. Optimally, this would be through encouraging publication venues to integrate lessons learned from negative results as part of their program. Other means, such as social media or reports on arXiv could also serve as a venue for bringing information about negative outcomes to the public. We also encourage the research community to organize workshops, Ph.D. forums, and other early-stage events for getting feedback and sharing information on how different researchers—also those in senior positions—have overcome the setbacks they have faced.

Communicating stories of the research process as a whole should be seen as equally important as communicating the final results. This could be a section at the end of the article, similar to the "limitations" section currently included in the discussion of the results. In addition, communicating the story of the work, from the initial hypothesis through failures to the final decision, should become a standard when discussing the results in conferences or sharing results on (research) social media, such as Twitter, ResearchGate, or LinkedIn. Here, encouraging senior researchers as role models to communicate about setbacks and failures—and how they did or did not overcome them—is crucial. Some could even consider publishing a list of rejections along with acceptances. Normalizing failure could even encourage researchers to take more risks in research without the fear of not having to show anything.

CONCLUSION

Failures and setbacks in research work are natural and will occur for everyone, from junior Ph.D. students to senior professors. In pervasive computing research, reasons for failures can become apparent quite late and might be unrecoverable, including overlooked aspects in data gathering or instrumentation, wrong assumptions, trust, and ethical considerations. Causes for failures can be common but relatively unknown as published research tends to highlight novel findings and overlook negative results, setbacks, and stories of difficulties. Sometimes writing and getting negative results published can be much more difficult than publishing a paper with minor incremental results.

Regarding the lesson learned from PerFail2022, a brief set of lessons for following studies in pervasive computing starts with careful analyses from previous studies in different aspects before deciding clear questions and suitable experiments with datasets as

well as the validity of the ground truth. Also, collecting data prioritizes high-quality and exact questionnaires detailing target groups and domain experts. Last, the stepwise procedures enable failure recognition during the study period.

As we have highlighted in this article, venues for discussing experiences from failures must become more mainstream. Shared experiences and lessons learned will help others to avoid failures; this should be done by presenting setbacks and talking to other members of the community starting from the early stages of the study. It would be exceptionally important to have a platform where people can publish the stories behind successful papers describing the problems they ran into. This could also be a specific section in each paper for failed trials, similar to how limitations are currently discussed. For our own part, we will continue running the PerFail workshop on negative results in pervasive computing. We encourage increasing the number of similar workshops also in other fields of computer science.

REFERENCES

1. N. Matosin, E. Frank, M. Engel, J. S. Lum, and K. A. Newell, "Negativity towards negative results: A discussion of the disconnect between scientific worth and scientific culture," *Dis. Models Mechanisms*, vol. 7, pp. 171–173, 2014.
2. A. Mlinarić, M. Horvat, and V. Šupak Smolčić, "Dealing with the positive publication bias: Why you should really publish your negative results," *Biochemia medica*, vol. 27, no. 3, pp. 447–452, 2017.
3. P. M. Dahlgren, "Instructions for reviewer 2: How to reject a manuscript for arbitrary reasons," May 2022. [Online]. Available: osf.io/t8jsm
4. S. Gaillard and S. Devine, "Systemic problems in academia: The positive publication bias and solutions from a human factors perspective," *J. Trial Error*, vol. 2, no. 1, pp. 1–5, 2022. [Online]. Available: https://assets. pubpub.org/ydsc2fci/11647096441542.pdf
5. S. Nimpf and D. A. Keays, "Why (and how) we should publish negative data," *EMBO Rep.*, vol. 21 2022, Art. no. e49775. Accessed: May 24, 2022, doi: 10.15252/embr.201949775.
6. A. Bespalov, T. Steckler, and P. Skolnick, "Be positive about negatives–recommendations for the publication of negative (or null) results," *Eur. Neuropsychopharmacol.*, vol. 29, no. 12, pp. 1312–1320, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924977X19317195
7. V. Bajpai et al., "The Dagstuhl beginners guide to reproducibility for experimental networking research," *SIGCOMM Comput. Commun. Rev.*, vol. 49, no. 1, pp. 24–30, Feb. 2019, doi: 10.1145/3314212.3314217.
8. L. Echevarria, and A. Malerba, and V. Arechavala-Gomeza, "Researcher's perceptions on publishing "negative" results and open access," *Nucleic Acid Therapeutics*, vol. 31, no. 3, pp. 185–189, 2021, Accessed: May 24, 2022.
9. M. de Vos, G. Ishmaev, J. Pouwelse, and S. Roos, "A deployment-first methodology to mechanism design and refinement in distributed systems," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events*, 2023.
10. D. L. Tennenhouse and D. J. Wetherall, "Towards an active network architecture," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 5, pp. 81–94, 2007.
11. M. Welzl, J. Ott, C. Perkins, S. Islam, and D. Kutscher, "How not to IETF: Lessons learned from failed standardization attempts," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events*, 2023.
12. N. Howell, A. Desjardins, and S. Fox, "Cracks in the success narrative: Rethinking failure in design research through a retrospective trioethnography," *ACM Trans. Comput.-Hum. Interact.*, vol. 28, no. 6, pp. 1–31, Nov. 2021, doi: 10.1145/3462447.
13. S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Front. Psychol.*, vol. 9, pp. 114–133, 2018, doi: 10.3389/fpsyg.2018.00861.
14. M. Heck, J. Edinger, and C. Becker, "Lessons learned from an eye tracking study for targeted advertising in the wild extraction," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events*, 2022, pp. 539–544.
15. S. Liaqat et al., "Hindsight is 20/20: Retrospective lessons for conducting longitudinal wearable sensing studies," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events*, 2022, pp. 545–550.

16. B. van Berlo, T. Ozcelebi, and N. Meratnia, "Insights on mini-batch alignment for WiFi-CSI data domain factor independent feature extraction," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events*, 2022, pp. 527–532.

17. A. Das, Y. D. Kwon, J. Chauhan, and C. Mascolo, "Enabling on-device smartphone GPU based training: Lessons learned," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events*, 2022, pp. 533–538.

18. A. S. Franzke, A. M. BechmannZimmer, and C. Ess, "Internet research: Ethical guidelines 3.0," Association of Internet Researchers, 2020.

19. D. R. Thomas, S. Pastrana, A. Hutchings, R. Clayton, and A. R. Beresford, "Ethical issues in research using datasets of illicit origin," *Proc. ACM SIGCOMM Internet Meas. Conf.*, 2017, pp. 445–462.

20. A. Borji, "Negative results in computer vision: A perspective," *Image Vis. Comput.*, vol. 69, pp. 1–8, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885617301671

ELLA PELTONEN is an assistant professor with the University of Oulu, 90014, Oulu, Finland. Her research focuses on pervasive computing, intelligent edge-cloud continuum, applied machine learning and artificial intelligence, and sensing data analytics in mobile and vehicular computing. She received her Ph.D. degree from the University of Helsinki. She is the corresponding author of this article. She is a member of IEEE. Contact her at ella.peltonen@oulu.fi.

NITINDER MOHAN is a senior researcher in the Chair of Connected Mobility at Technical University of Munich, 00014, Munchen, Germany. His research interests include edge computing, next-generation networked applications and large-scale Internet measurements. He received his Ph.D. (as MarieCurie ITN fellow) degree from the Department of Computer Science, University of Helsinki, Finland. Contact him at mohan@in.tum.de.

PETER ZDANKIN is currently working toward the Ph.D. degree with the University of Duisburg-Essen, D-45117, Duisburg, Germany. His current research interests include novel IoT architectures focusing on longevity in various aspects. Contact him at peter.zdankin@uni-due.de.

TANYA SHREEDHAR is a postdoctoral researcher with the University of Edinburgh, EH8 9YL, Edinburgh, U.K. Her research interests lie broadly in the area of networks and systems with a focus on transport protocols and data-driven networking. She received her Ph.D. degree from IIIT-Delhi. Contact her at tanya.shreedhar@ed.ac.uk.

TRI NGUYEN is a doctoral researcher at the Center for Ubiquitous Computing, University of Oulu, 90014, Oulu, Finland. His research interests include distributed systems, blockchain technology, and information security. His current research is about blockchain applications in transportation and federated learning and the scalability of blockchain. He received his M.S. degree in computer science from the University of Pisa, Italy. Contact him at tri.nguyen@oulu.fi.

SUZAN BAYHAN is an assistant professor at the University of Twente (UT), 7522 NB, Enschede, The Netherlands. Her research interests include energy efficiency and resilience of wireless communication systems, resource management for cellular networks, and edge computing. She received her Ph.D. degree in computer engineering from Bogazici University. Contact her at s.bayhan@utwente.nl.

JON CROWCROFT has been the marconi professor of communications systems in the Computer Laboratory since October 2001 in University of Cambridge, CB2 1TN, Cambridge, U.K. His three main topics of interest have been scalable multicast routing, practical approaches to traffic management, and the design of deployable end-to-end protocols. His current active research areas are opportunistic communications, social networks, privacy preserving analytics, and techniques and algorithms to scale infrastructure-free mobile systems. He is a fellow of IEEE. Contact him at jon.crowcroft@cl.cam.ac.uk.

JUSSI KANGASHARJU is a full professor and the head of the department of Computer Science, University of Helsinki, 00100, Helsinki, Finland. His research interests are information-centric networks, edge and cloud computing, content distribution, opportunistic networks, and green ICT. He received his Ph.D. degree from the University of Nice Sophia Antipolis/Institut Eurecom in 2002. He is a member of IEEE and ACM. Contact him at jussi.kangasharju@helsinki.fi.

DANIELA NICKLAS is a full professor with the University of Bamberg, 96047, Bamberg, Germany, and holds the Chair of Computer Science, in particular Mobile Software Systems/Mobility. Her research interests include computer systems that bridge the gap between the physical world and the digital world. Contact her at daniela.nicklas@uni-bamberg.de.