A Study on Reproducibility in Computational Creativity Research

Martin Marzidovšek
Jožef Stefan International Postgraduate School
Jožef Stefan Institute
1000 Ljubljana, Slovenia
martin.marzidovsek@ijs.si

Senja Pollak
Jožef Stefan Institute
1000 Ljubljana, Slovenia
senja.pollak@ijs.si

The importance of reproducibility has been recognised as one of the crucial elements of research especially in life sciences, but also in other disciplines, e.g. medicine, social sciences, natural language processing who have acknowledged the reproducibility crisis. In this paper, we discuss the issue of reproducibility in the field of computational creativity. We present the findings from an indirect reproducibility study that assesses the transparency of available information to allow reproducibility. It does so through the analysis of articles published in ICCC proceedings, which has shown that most computational creativity (CC) publications until now have provided very little with regard to documentation and resources needed for reproducibility and replicability of the published results. By reviewing best practices from the broader scientific community and considering the particularities of CC recommendations are put forward that will hopefully inspire the creation of standards and practices related to reproducibility and improve the reusability of CC research within the CC community and beyond.

Introduction

The goal of science is acquiring knowledge about the world for which it depends on the "ability of the scientific community to scrutinize scientific claims and to gain confidence over time in results and inferences that have stood up to repeated testing" (CRRS[1] 2019). Science is thus an iterative but also collaborative process that advances (more efficiently) when researchers can build upon others' work, reproduce and reuse their results. For this to work scientists should as much as possible share data, methods and results, report uncertainties about their findings, and share not only positive, but also negative results.

While it is hard to claim that science is fully "objective", building trust in science requires standards and procedures that are accepted by scientific community. Karl R. Popper (1934 [2002]) claimed that "the objectivity of scientific statements lies in the fact that they can be inter-subjectively tested". So for him "objectivity" did not depend so much on the correspondence of a scientific claim to facts as it did on its verifiability: whether it can be tested and put under rational scrutiny. Over the years, common scientific practices and standards have evolved such as *reproducibility*, which many see as a cornerstone of the scientific method since it enables the scientific community to confirm or refute research results, but also allow for reusability of prior research.

Reproducibility is not by itself concerned with the correctness of the results or the process. More importantly, as long as research is reproducible bugs in the code and flawed methodology can become transparent for other researchers who can improve the original work and so scientific critique and progress can be made. Reproducibility thus stands for providing a complete and unambiguous description of the entire process from raw data to the final results. "[…] When a researcher transparently reports a study and makes available the underlying digital artefacts, such as data and code, the results should be computationally reproducible" (CRRS 2019).

While *reproducibility* is concerned with obtaining quantitative scientific results by independent scientists using the original datasets and methods, *replicability* is as important, since it concerns validation of specific findings with other datasets and implementations of the original methods (Stodden 2014, Branco et al. 2017). Not less important is *reusability* referring to the capacity to reutilize a novel component from the original research in another system even when insufficient resources are provided to allow for reproduction of experiments.

In the recent years, however, concerns about reproducibility and replicability – widely considered as hallmarks of good science – have increased in the scientific community and critical articles in high-profile mainstream media even spoke of a reproducibility "crisis" (Nature 2016, The Economist 2013, Branco et al. 2017). The quality of empirical results has been questioned and importance of reproducibility highlighted in many fields, such as (bio)medicine (e.g. Prinz et al. 2011; Ioannidis 2005), neuroscience (Button et

[1] Committee on Reproducibility and Replicability in Science (US National Science Foundation).

Proceedings of the 11th International
Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

al. 2013), economics (Camerer et al. 2016), language technologies (Branco et al. 2017) etc.

The ability and effort required from other researchers to replicate experiments and explore variations depends heavily on the information provided when the original work was published (Gundersen et al. 2018). We believe that, even if not applicable to all types of papers, for majority of scientific papers which describe empirical studies one can claim that if published research is not reproducible, it is of much less value.

By taking into account the specificities of Computational Creativity (CC) field this paper elaborates on the importance for the CC community to share with their published research complete and sufficient documentation about the used digital artefacts (e.g. datasets, code), methods and complete results to facilitate reproduction and replication.

Relevance of Reproducibility for the CC Field

As a sub-field of Artificial Intelligence research, Computational Creativity (CC) is the philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative. (Wiggins, 2006; Colton and Wiggins, 2012). As mentioned by Colton and Wiggins (2012) the methodological requirements for evaluation are crucial for the field, where they emphasise the involvement of unbiased observers in fairly judging the behaviours exhibited by CC systems.

While in CC there are different types of papers and contribution made, we believe that at least for technical and system presentation papers, the reproducibility is very important underlying aspect, which would allow for better and quicker development of novel systems in the field, more transparent evaluation of different systems, easier update of the research by novel researchers in the field, and better integration of previous work in novel systems. One could question, whether in the field where creativity and novelty is at the core, there is place for reproducibility, but we argue that striving for scientific research standards could be beneficial to an evolving field such as CC and could help its progress. As has been shown already by Platt (1964) particular scientific fields move more rapidly because they adopt systematic research methods.

Thus, in similar ways as other scientific fields, CC could try to rely more on reproducible experiments to validate research results, new discoveries and practices. We propose that the CC community should, whenever possible, strive to facilitate reproducible and replicable research by adequate experimental design and methods as well as clear and complete documentation in the publications. We advise to establish practical and pragmatic practices on how to document the scientific methods and resources so that reproducibility and replicability of CC research results is feasible in practice. In addition, sharing the code and developed tools would allow easier introduction to the field of new researchers and would be beneficial for the promotion of the field across disciplines and communities.

This paper presents the findings from an indirect reproducibility study that assesses the transparency of available information in published CC research to allow reproducibility. As will be shown in the analysis of International Conference on Computational Creativity (ICCC) proceedings most CC publications until now have provided very little in regard to documentation and resources that would allow, let alone facilitate, reproducibility and replicability.

It is crucial for CC as a scientific field to understand the reasons for that and to address these issues with the goal of improving transparency of the published research. While this paper can't fully answer this question it aims to raise awareness of these issues among the CC community and inspire further inquiries and work that would bring about better reproducibility and replicability practices that would subsequently lead to fewer studies that do not reproduce or replicate.

Background and Related Work

In response to the already mentioned concerns about reproducibility of science by both the scientific and mainstream media the US Congress initiated in 2017 an assessment that resulted in a comprehensive 2019 report *Reproducibility and Replicability in Science* prepared by the Committee on Reproducibility and Replicability in Science (CRRS). Because the terms *reproducibility* and *replicability* are sometimes used interchangeably or have different and even conflicting meanings depending on the scientific field we use the definitions proposed by CRRS (2019):

> *"Reproducibility* is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis."
>
> *"Replicability* is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data."

The same Committee further specifies that reproducibility involves the original data and code, while replicability involves new data collection to test for consistency with previous results of a similar study.

From a slightly different angle: "replicability or repeatability is a property of an experiment: the ability to repeat – or not – the experiment described in a study" (Cohen et al. 2018). Reproducibility, on the other hand, is a "property of the outcomes of an experiment: arriving – or not at the same conclusions, findings, or values" (ibid.).

In addition to the above two terms we point also to *reusability* of the code. So when complete experimental resources necessary for reproduction or replication are not available, we believe that reuse of the accessible components can still be beneficial as it allows the community to collaborate and develop further as well as trigger more interest from the neighbouring scientific and engineering domains.

Results from Reproducibility Studies

As noted by Peng (2016), reproducibility which initially sounds like a trivial task has shown that it's not always easy to achieve. The CRRS reviewed a collection of reproducibility studies across a variety of scientific fields and found that "[…] systematic efforts to reproduce have failed in more than one-half of attempts made, mainly due to insufficient detail on digital artifacts, such as data, code, and computational workflow" (CRRS 2019).

The famous 2016 survey "Is There a Reproducibility Crisis?" by Nature assessed ten factors that turned out to majorly contribute to irreproducible research and came to similar conclusions. Among the factors were: selective reporting, methods and code unavailable, poor experimental design, pressure to publish, low statistical power, not replicated enough in original lab, raw data not available from original lab, and others (Nature 2016).

A study on reproducibility in artificial intelligence (AI) sampled 400 papers from the AAAI and IJCAI conferences and found that computational AI research was not documented systematically and with enough information to support reproducibility (Gundersen and Kjensmo 2018).

Similarly in a recent study Repar et al. (2019) analysed several influential papers on bilingual terminology extraction (a field of NLP) from the past 25 years where they assessed the dataset, code and tool availability for the purpose of reproducibility and replicability. A surprising observation was that not one from the sampled papers made experiment code available and only a few provided links to tools where experiments were conducted. This severely hinders replicability. Repar et al. furthermore attempted to replicate one of the analysed papers (Aker et al. 2013) and despite closely following the original paper they obtained significantly worse results than the paper's authors.

Sources of Irreproducibility and Irreplicability

The above studies exemplify how difficult of a task it is to reproduce or replicate research that lacks experiment code, datasets and complete information about the implementation. The CRRS (2019) also claims that the greatest barriers to reproducibility are inadequate recordkeeping and non-transparent reporting. It thus follows that efforts to encourage more transparency in scientific publications would be beneficial.

According to CRRS (2019) there are a number of factors that make reproducibility of published research so difficult to achieve. In addition to missing access to non-public data and code the Committee also mentions inadequate record keeping (steps followed), non-transparent reporting, obsolescence of the digital artefacts, etc.

There is an important "conflict" of incentives between researchers who conducted the initial study and the independent researchers who attempt to reproduce the results. As Gunberrger et al. (2017) have argued: "independent researchers trust an empirical study's results increasingly with the amount of documentation that is shared with them, while the effort to reproduce the results increases when the amount of documentation is reduced. […] On the other hand, the effort to document the research increases for the original researchers with the amount of documentation that needs to be shared, while the generality of the method is increased if independent researchers reproduce the results given less documentation".

All research should be reproducible but it can be expected that not all research will replicate due the inherent risks of statistical procedures, researchers' mistakes and biases (e.g. selection of methods that confirm the desired hypothesis, or selecting the hypothesis only after seeing the data, splitting data into subsets that lead to desired results, etc.). Consequentially preregistration is becoming more and more an accepted norm to deal with these problems. It requires researchers to register all relevant aspects and information about the scientific study (data collection, hypotheses, methods used, etc.) before they start with the research.

Reproducibility and Replicability Best Practices

The CRRS issued several recommendations for ways on how researchers, academic institutions, journals, and funders should help strengthen rigor and transparency in order to improve the reproducibility and replicability of scientific research. One of the most important recommendations for our analysis states (recommendation 4-1): "*To help ensure the reproducibility of computational results, researchers should convey clear, specific, and complete information about any computational methods and data products that support their published results in order to enable other researchers to repeat the analysis, unless such information is restricted by non-public data policies. That information should include the data, study methods, and computational environment"* (CRRS 2019).

As proposed by Repar et al. (2019) availability of datasets is an essential prerequisite for successful replication, while having access to original code greatly increases the ease or reproducibility and replicability experiments. In terms of record keeping a full compendium of artefacts is required from the original researcher. The CRRS specifies that the "computational details that need to be captured and shared for reproducible research include data, code, parameters, computational environment, and computational workflow" (CRRS 2019).

The CCRS report states that even when a project's data are publicly available the analytical methods described by authors in scientific papers often lack sufficient guidance to reproduce the results. As also noted by Repar et al. "[...] even code itself is sometimes not enough without additional implementation notes and information on the operating systems and software used." The recommendation 6-1 by the CRRS (2019) addresses this problem: "*All researchers should include clear, specific, and complete description of how the reported result was reached.*" The Committee further specifies, which details should be included (e.g. clear description of all methods used, data management, discussions of uncertainty, etc.).

The modern standard for reproducibility of research is to use computer code for everything. Meaning, it should be

Proceedings of the 11th International
Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

avoided to do even minimal changes to the data (pre-processing), generation of visualisations and so on manually as it can introduce errors and such workflows are mostly not recorded and thus undetectable to others. Computer code, on the other hand is way less ambiguous and there is less room for misinterpretation.

Of course not all researchers have high computer programming skills and there are other ways how they can share or execute reproduction and replication experiments. Instead of, or better, in addition to sharing experiment source code a web tool (e.g. ClowdFlows[2]) can be used to replicate the experiment and enable others – especially those with less or no programming skills – to reproduce their results. "Availability of a tool or application (online or offline) where experiments can be conducted eases reproducibility and replicability, but also enables the reusability of results by a larger community" (Repar et al. 2019).

Journals and scientific societies have an especially important role to play in improving the reproducibility of research output. As is suggested in recommendation 6-7 CRRS (2019): "*Journals and scientific societies requesting submissions for conferences should disclose their policies relevant to achieving reproducibility and replicability*." The CRRS encourages these entities to set and implement desired standards of reproducibility and replicability and make this one of their priorities. It also proposes the adoption of policies to reduce the likelihood of non-replicability with specific measures.

At Nature authors when "submitting manuscripts to Nature journals would need to complete a checklist addressing key factors underlying irreproducibility for reviewers and editors to assess during peer review […]. Nature's checklist was designed, in part, to make selective reporting more transparent. Authors are asked to state whether experimental findings have been replicated in the laboratory, whether and how they calculated appropriate sample size" (Nature 2018a).

Gundersen et al. (2017) points out that reproducibility is concomitant with open science, which involves sharing data, software, and other science resources in public repositories using permissive licenses. He also notes the FAIR Guiding Principles for Scientific Data Management and Stewardship, which are increasingly associated with open science and ensure that science resources have the necessary metadata to make them findable, accessible, interoperable, and reusable (Wilkinson et al. 2016).

Connected to open sicence is also modern digital scholarship that promotes credit to scientists who document and share their research products through citations of datasets, software, and innovative contributions to the scientific enterprise (Gundersen et al. 2017).

Reproducibility in CC Community

CC can be seen as a field of science with diverse research and engineering methods and output. It is thus not always appropriate to draw direct comparisons with other scientific research areas. Best practices from physics or life sciences are not necessarily transferable as the results and evaluations of scientific inquiries are more uniform and adapted to that particular field.

Nonetheless, CC should as far as appropriate compare its reproducibility and replicability practices to standards established in the general scientific domain and best practices from similar disciplines such as artificial intelligence. This is crucial so CC can stay in touch with reproducibility trends and will drive it to maintain the highest possible scientific standards so it can strengthen its status as a scientific discipline.

Computational creativity research as it is now defined and centred around the Association for Computational Creativity is still a relatively new field. In 1999 the yearly International Joint Workshops on Computational Creativity started but an even wider global research community began to evolve with ICCC conferences the first of which was held in Lisbon 2010. With an international conference and increased research activity CC as a field of research started to mature and establish its place in the broader scientific community.

As has been explained earlier the cornerstone of science is verification – the process by which scientists confirm the validity of new findings or discoveries by repeating the research that produced it (CRRS 2019). This has lead us to reproducibility and replicability as practices ensuring that this is possible and our central goal of this paper – to review these practices among the CC community.

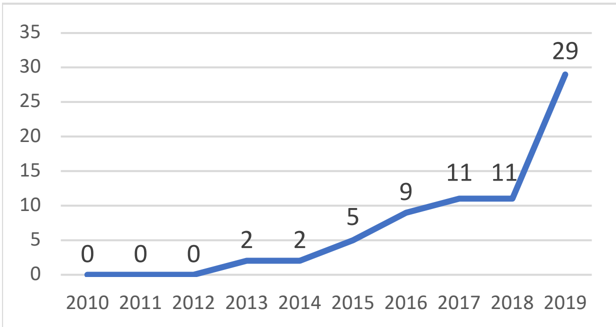
To get a bird's-eye-view about the state of reproducibility and replicability of the published research at the past ICCC conferences we analysed what we consider to be the centrepiece of modern reproducibility and replicability practice – the sharing of experiment source code, datasets and other relevant resources for the published research by means of public online repositories. Nowadays, there is a plethora of repositories for sharing datasets (see e.g. CLARIN initiative), and code (e.g. GitHub, GitLab).

We decided to measure GitHub links as a simple indicator to give an approximation about the trend and frequency of referencing digital repositories in CC articles published in ten years of ICCC proceedings (2010-2019). While acknowledging that GitHub by no means is the only possible way to share reproducibility related documentation it has since its launch in 2008 become the largest host of source code in the world and is thus a good candidate to give a hint on the practice of sharing digital artefacts such as source code and datasets in digital repositories within the CC community.

As can be seen in Graph 1, the GitHub links only slowly begin to appear in 2015, which is the year when GitHub already became the largest public online repository in the world, and a relative increase can be observed in 2019. However, as will be shown in the more detailed analysis of

[2] http://clowdflows.org

the ICCC proceedings of 2019 the actual links hosting resources necessary to reproduce the authors' published research are even in 2019 still quite rare – many of the included GitHub links point to third party resources used such as libraries, related source code, etc.



Graph 1: Trendline of GitHub Links in ICCC Proceedings 2010-2019.

Analysis

In order to get a better understanding about where the CC field is now with regard to reproducibility and replicability of research, we decided to analyse closer the CC articles published in the Proceedings of ICCC 2019 (Grace et al. 2019), which offers a good overview on where the field currently stands. It is important to note that the conference in 2019 marked the 10th anniversary of ICCCs, which signals that the CC as a field of research is maturing and that research practices and publishing standards in this scientific community are by now sufficiently established.

We have selected 34 articles from the proceeding (ibid.) that fall in the category of *technical* and *system and resource description*. The identification of the category has not been clear-cut as the category is not stated in the proceedings. The sample tried to include all of the articles that use empirical research or engineering approaches, where verification and thus reproducibility and replicability should be possible. We have omitted all of the articles from the *creative submissions*, those proposing theoretical frameworks, and methodologies.

As there are no standards yet regarding reproducibility and replicability in the CC community we relied on best practices in the wider scientific community presented earlier. The analysis of the sampled papers does not intend to expose, which tick-boxes a certain article failed to satisfy but rather to give an indication on which of the important digital artefacts are currently being shared in the CC community's research papers.

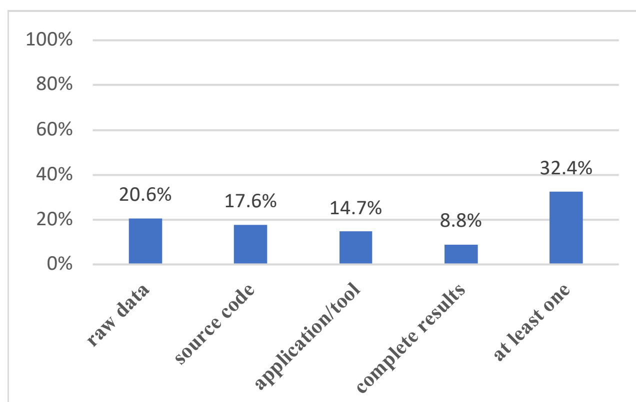
The selected articles have been analyses for four indicators: *raw data*, *source code*, *application or online tool*, and *complete results*. These represent the minimum standards for reproducibility and replicability of scientific research that we consider crucial for the progress of CC as a field of scientific research. They are also easy to identify and are likely be considered indisputably beneficial for the progress of open science and research by the CC community. However, deeper analysis of the factors relevant for reproducibility residing in the content of the articles is needed once more clear standards and policies by the ACC are established.

In the four selected indicators we were looking for whether the particular digital artefact has been shared by the author via a *link* directing to a digital repository or website that included complete or sufficient resources that would allow reproduction and replication of the study. This is in line with the already mentioned recommendation by CRRS (2019) stating that "researchers should convey clear, specific and complete information about any computational methods. […] That information should include the data, study methods, and computational environment."

No special software has been used for identifying the selected indicators, which have been searched and analysed manually from the pdf version of Proceedings of ICCC 2019 (the same approach has been used to count GitHub links in ICCC proceedings (2010-2019)).

Findings

The analysis of the 34 sample papers from the ICCC 2019 proceedings (Grace et al. 2019) shows that only a minority of articles that have been accepted to the ICCC 2019 include the most crucial elements needed for reproducibility and replicability. As can be seen on Graph 2 only 20.6% of the analysed articles shared the dataset, 17.6% shared the source code, 14.7% provided links to web tools or applications, and only 8.8% shared complete results.



Graph 2: Percentage of Papers Sharing Reproducibility Resources (Indicators).

Only 32.4% of papers provided at least one of the analysed digital artefacts, which means that more than two thirds of the published papers did not provided any kind of resource that is indispensable for the reproduction or replication of the published research. This raises the question why authors don't provide access to artefacts related to their research: is it a lack of concern, time, or will for the extra

Proceedings of the 11th International
Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

effort needed to prepare the most basic resources or an intentional decision of keeping the research for themselves for various reasons?

While most papers offer at least some sort of description and sometimes mention the source of the dataset used for their experiments this does not suffice to allow others to reproduce their work. Also providing links to home pages from where the data has been obtained is not really useful for replication as the exact dataset would be needed.

A few papers mentioned or provided links to third party code that has been used or adapted for their research (e.g. libraries of specific components) this did not match the requirements of our source code indicator as only the complete experiment code that was used for the published research allows it to be reproduced and replicated.

In general it seems to be the prevalent praxis that authors in their papers include only samples of the data, code, pseudocode, results or screenshots from their web applications. This does not permit others to reproduce or replicate the results and to allow new members of the community or researchers from other scientific communities to be able to develop upon previous research.

One of the few provided links to web applications presented in some of the papers and one link to results were broken. In the first case we counted the link as being provided as a working link could be found in the shared GitHub repository. As more links will become obsolete with time ways to make access to resources permanent should be sought.

There have also been cases where functioning web platforms that resulted from the research are mentioned but no links are provided. If the research output is proprietary and the source code or the applications built are not intended for the public use it would be appropriate to clearly state this.

Discussion and Recommendations

We recognise that the predominant interest in CC is for the final research output (e.g. creative systems and artifacts) and that because of this documentation of other aspects of the CC research can fall short, which could on the long-term have negative effects for the CC field. It is thus important to draw attention to the described practices of reproducibility and replicability of scientific research that will allow CC to develop more efficiently.

CC researchers should follow best practices from other scientific disciplines and "describe methods and data in a clear, accurate and complete way" (CRRS 2019). In addition accessibility of complete results is especially important for the now predominant cases where source code and tools are not available together with the input dataset. As has been argued by Colton and Wiggins (2012), "in many projects, the output is carefully scrutinised by the program's author, and only the best examples are shown to audiences, or used as exemplars in research papers, etc." They refer to the *curation coefficient* as a means to understand the performance of a particular creative system. In this regard the availability of complete results give a better sense on the representability of the few output examples that are usually included in the papers.

Publishing and research practices that were discussed earlier and are presented in our recommendations can improve reproducibility and replicability in the CC research but they clearly require additional efforts from the authors particularly as they incorporate them in their work habits. But there are also direct benefits to the authors. Gundersen et al. (2017) state ten benefits, among them: practice open science and reproducible research; receiving credit for all your research products (by citing software, datasets, and other products); increase the number of citations to your publications (well-documented articles receive more citations); improved chances of being funded; improved management of your research assets, etc.

The benefits to the CC community are clear: maintaining repositories for CC code, tools and data, would make the field more attractive to young researchers, facilitate teaching of CC and allow for fruitful exchange of results between the neighbouring fields (e.g. NLP and CC).

What follows is a list of some general recommendations relating to data repositories, source code, the presentation of results, and more that could inspire standards for facilitating reproducibility.

Reproducibility and Replicability Recommendations for the CC Community (Researchers and Institutions):

1. The complete data, source code and results should be findable and accessible via shared open repositories (e.g. GitHub, GitLab, etc.); while informative, samples of the data and source code, pseudocode, tool screenshots, etc. in the paper are not enough.
2. In case of dataset restrictions (e.g. privacy, intellectual property constrains) or other valid reasons for not sharing digital resources, this should be explained. When feasible, permissions should be arranged for reproducibility purposes.
3. Provide description and documentation of the experimental design, hardware and software used, and a digital record of the workflow (e.g. data selection and manipulation, parameters, results at different stages, etc.). Use computer code for every step in the experiment as it provides a clearer and less unambiguous record of all steps (e.g. avoid manual data manipulation that can introduce errors and is undetectable).
4. Use integrated analyses and reports that combine code with data manipulation and visualisations (e.g. Jupyter notebook).
5. Strive for at least one replication by e.g. colleagues before submission for publication, which will expose weaknesses in the documentation.
6. ACC should promote clear transparency requirements through standards and policies for reproducibility and replicability.

7. ICCCs could introduce a reproducibility and replicability submission checklist.
8. ACC could promote and publish more replication studies (e.g. as a special paper type in the ICCC call for papers).
9. ACC could promote and support preregistration before the start of CC research to avoid replicability pitfalls (e.g. researcher's biases, etc.).
10. Make proper use of statistical methods; estimate and report uncertainties in results.
11. Use persistent links (PURL) for all shared resources or have associated digital object identifiers (DOI) so the shared resource is findable and available permanently.
12. ACC could establish an open online repository for research output submitted to ICCCs so it would be accessible on one place for the CC community and other interested parties.
13. The ACC and ICCC's should promote the use of web tools in research (e.g. ClowdFlows) that would make reproducibility simpler and bring CC research results closer to people with less programming skills.
14. All members of the CC community should as much as possible promote, practice and support open science, the FAIR principles and digital scholarship.

Other relevant recommendations for data, source code, experiments and workflows, digital records and more can be found among others in Gundersen et al. (2017), CRRS (2019) and Repar et al. (2019).

Conclusions and future work

We have tried to argue that the CC community needs to recognize the benefits of reproducible science. We believe the CC community should strive for a more open and collaborative research by sharing, verifying and building upon each other's research. Because of this CC researchers, publishers and institutions should strive to adopt better reproducibility and replicability practices and ensure that its research is transparent and well documented to make reproducibility as feasible as possible in practice.

The present analysis of the factors for reproducibility and replicability in CC research has only looked at the most basic requirements. As potential future studies might show any attempt at reproducing CC research in practice will likely need more detailed information and digital record of the entire experiment: data manipulation, parameter settings, workflow, software, interim and complete final results, etc.

The findings of this indirect reproducibility study open the question on how the necessity to reproduce and replicate scientific results is perceived by the CC community. What are the reasons for the relatively low attention given to these aspects?

In future work, we plan to analyse a larger sample of the archives, and complement the research with a survey and interviews, which would allow for understanding of the researchers' attitudes towards this topic. We strongly believe that encouraging the community to enable reproducible experiments by sharing the resources and tools is of crucial importance and can drive CC as a scientific field even further. It will increase visibility and position it along other scientific strands that build upon the foundations of falsifiable and verifiable science that allows human knowledge to expand. Isn't this the goal after all?

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research Agency for research core funding for the programme Knowledge Technologies (No. P2-0103).

References

Aker, A., Paramita, M.L. and Gaizauskas, R., 2013. Extracting bilingual terminologies from comparable corpora. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 402-411.

Baker, M., "Reproducibility crisis?". *Nature,* 2016. Available: https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970 [March 2020]

Branco, A., Cohen, K.B., Vossen, P., Ide, N. and Calzolari, N., 2017. *Replicability and reproducibility of research results for human language technology: Introducing an LRE special section*.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S. and Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), pp.365-376. Available: https://www.nature.com/articles/nrn3475 [July 2020].

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T. and Heikensten, E., 2016. Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), pp.1433-1436.

Cohen, K.B., Xia, J., Zweigenbaum, P., Callahan, T.J., Hargraves, O., Goss, F., Ide, N., Névéol, A., Grouin, C. and Hunter, L.E., 2018. Three dimensions of reproducibility in natural language processing. In *LREC... International Conference on Language Resources & Evaluation:[proceedings*]. *International Conference on Language Resources and Evaluation (Vol. 2018)*, 156. NIH Public Access.

Colton, S. and Wiggins, G.A., 2012. Computational creativity: The final frontier?. In *Ecai (Vol. 12)*, 21-26.

[CRRS] Reproducibility and Replicability in Science, 2019.

National Academies of Sciences, Engineering, and Medicine. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303.

How Science Goes Wrong, 2013. *The Economist* (Oct 21 2013). Available: https://www.economist.com/leaders/2013/10/21/how-science-goes-wrong [March 2020].

Grace, K., Cook, M., Ventura, D. and Maher, M. L., 2019. *Proceedings of the 10th International Conference on Computational Creativity*. Association for Computational Creativity (ACC).

Gundersen, O.E., Gil, Y. and Aha, D.W., 2017. On reproducible AI: Towards reproducible research, open science, and digital scholarship in AI publications. In *AI magazine (39(3))*, 56-68.

Gundersen, O.E. and Kjensmo, S., 2018. State of the art: Reproducibility in artificial intelligence. In *Thirty-second AAAI conference on artificial intelligence*.

Ioannidis, J.P., 2005. Why most published research findings are false. *PLoS medicine*, *2*(8), p.e124. Available: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124 [July 2020].

"Checklists work to improve science." *Nature*, 2018a. Available: https://www.nature.com/articles/d41586-018-04590-7 [March 2020]. doi: 10.1038/d41586-018-04590-7.

"Challenges in irreproducible research." *Nature,* 2018b. Available: https://www.nature.com/collections/prbfkwmwvz [March 2020].

Peng, R.D., 2016. A Simple Explanation for the Replication Crisis in Science. Available: https://simplystatistics.org/2016/08/24/replication-crisis

[March 2020].

Platt, J.R., 1964. Strong inference. *Science*, 146(3642), pp. 347-353.

Popper, K., 1934 [2002], *Logik der Forschung*, Akademie Verlag. English translation as *The Logic of Scientific Discovery*, Routledge.

Prinz et al. 2011, Believe it or not: how much can we rely on published data on potential drug targets?". *Nature,* 2011. Available: https://www.nature.com/articles/nrd3439-c1 [July 2020].

Repar, A., Martinc, M. and Pollak, S., 2019. Reproduction, replication, analysis and adaptation of a term alignment approach. *Language Resources and Evaluation*, 1-34.

Stodden, V., Leisch, F. and Peng, R.D. eds., 2014. *Implementing reproducible research*. CRC Press.

Wiggins, G. A. 2006. A preliminary framework for description, analysis and comparison of creative systems. Journal of Knowledge Based Systems 19(7):449–458.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3.