Data accessibility is not sufficient for making replication studies a matter of course

Huschka, Denis; Wagner, Gert G.

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
SSG Sozialwissenschaften, USB Köln

Empfohlene Zitierung / Suggested Citation:
Huschka, D., & Wagner, G. G. (2012). *Data accessibility is not sufficient for making replication studies a matter of course.* (RatSWD Working Paper Series, 195). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). https://nbn-resolving.org/urn:nbn:de:0168-ssoar-427921

RatSWD Working Paper Series

www.ratswd.de

RatSWD.
German Data Forum

195

Data accessibility is not sufficient for making replication studies a matter of course

Denis Huschka and Gert G. Wagner

April 2012

Federal Ministry of Education and Research

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

Contact: S C I V E R O Press | German Data Forum (RatSWD) | Mohrenstr. 58 | 10117 Berlin | office@ratswd.de

Data accessibility is not sufficient for making replication studies a matter of course

Denis Huschka und Gert G. Wagner

German Data Forum (RatSWD)

April 2012

Which code of behavior should form the basis of science and research? Replicability is definitely among these values. It is a pivotal feature of good scientific practice. Only replicable results are indeed scientific results. Studies that cannot be replicated are, strictly speaking, not scientific, but – given they are good – a type of feuilleton. Still, to most researchers – and this might seem surprising – facilitating and particularly conducting a replication study is anything but a matter of course.

In the social, behavioral, and economic sciences, few people replicate and review results of others despite the many data sets that, supposedly, can be accessed freely and analyzed by hundreds of scientists on an international scale. One exception are time series analyses based on the national economic accounts: here, the observation points are so scarce that replication and the improvement of research results by applying new methods, are necessarily part of the academic routine.

With that said, a recent survey by the multidisciplinary journal *Science* is particularly interesting. In December 2011, now for the second time, the journal dedicated a special to the issue of research data. Titled "Data Replication and Reproducibility", it discusses the various ways and means of replication of empirical studies (http://www.sciencemag .org/site/special/data-rep/).[1]

1 Empirical Results

The fact that replicability of empirical studies is an important scientific good, which is not just "produced" naturally, but in fact, has to be supported and acknowledged by the means of different mechanisms, was clearly conveyed by an online poll conducted by *Science* in the context of their "Data Replication and Reproducibility" special (cf. appendix). They asked the following question:

„Ideally, scientists would fully disclose their own raw data and methods and also spend time replicating others' work. What would best ensure this good behavior?"

The poll yielded the following results:

- Recognition and rewards from institutions : 6 %
- Funding earmarked for replication studies: 19 %
- More publication by journals of data that confirm or refute previous work: 45 %
- Rewards from funders on subsequent grant applications for depositing sufficient details for replication (or penalties for noncompliance): 24 %

[1] See also Siri Carpenter, Psychology's Bold Initiative: In an unusual attempt at scientific self-examination, psychology researchers are scrutinizing their field's reproducibili**ty,** *Science***,** Vol. 335 (30 March 2012) 1558-1561**.**

These results are interesting regarding theory and sociology of science. The greater part of the respondents – it can be surmised that most of them are scientists or researchers – demand tangible incentives that reward something that should be generally considered good scientific conduct, namely "reproducibility". This is a questionable stance – it seems, above all, scientific conduct has to be worthwhile: through material gains (funding) and –above all– reputational gain (publications).

The noble aspiration of scientific objectivity and reproducibility as such, seems to fall far short. Of course, the lack of understanding for the intrinsic value of scientific objectivity and reproducibility cannot be blamed simply on scientists alone. It appears the modern science system is characterized by a contradiction between the values of "textbook scientific conduct" and the opposing reality, which is characterized by inherent expediency and reward structures. Academics, especially young academics, who have to work hard (and fast) to gain reputation and to consolidate it, are subject to similar incentive structures worldwide: Publish or Perish – read: publish a multitude of publications in particularly renowned journals, or perish. Placement in these top journals is normally not achieved with replication studies. As important as these are for science and the preservation of the science system as a whole, replication of other peoples' work is a merit hardly rewarded. It usually requires solid evidence of plagiarism or grave methodological errors before an actual attempt is made to replicate publicized results. More commonly, different studies are compared and critically evaluated for "meta-analyses". Meta-analyses are widely considered distinct academic achievements, while replication studies are just considered boring. Meta-studies can expose attempts of plagiarism and poorly executed research, but not necessarily so. Instead, peculiar results are condoned as "rogue results".

This structural set-up affects behavior of empirical researchers: If, as in many cases, there is limited demand for my research data and procedure of analysis for replication, I will invest little time into editing my work and making it accessible for replication. And if, there are no prizes to be won with mere replication, no one will invest the effort. This is inbuilt in the system – and thus fully rational.

The situation in the social and economic sciences make the problem abundantly clear: these disciplines have been well-equipped, for decades in some cases, with data archives and interesting data sets that would facilitate replication studies in a comparatively simple way. Of course, hardly anybody does recalculations – in the literal sense. The insufficient replication of results despite widespread data availability in the social and economic sciences, underscore the following claim: general and easy availability of data does is not sufficient to encourage replication of results, which would contribute to ensuring quality in academic research. Individual incentives have to be effective, too.

2 Implications for science theory and organization

The *Science* poll on "Data Replication and Reproducibility" shows: Scientists and researchers from all disciplines expect rewards for matters of course, namely good scientific practice regarding methods and results. More (funding) money and more publications. Only 6% are content with simple „recognition".

The results do not show that academics are professionally unethical. But the expediency structures in the science system are clearly reflected in the answers. And they call attention to the currency that pays for better practice by means of systematic replication: publications. A long list of publications is essential to the recruitment process for senior positions in the science system, as opposed to the mere proof of adherence to the rules of good practice. That is taken for granted – until it is proved otherwise. Increasingly, this proof is provided: prominent scandals involving plagiarism have been mounting in recent years. Especially, so it seems, in the life sciences.

Lists of publication and the resulting assessment of citation measures for individual researchers are attempts to objectify academic achievements and to make them enumerable. To readers, the prominent title of a renowned journal is an indicator for relevance and for coherent results. This is a rather doubtful assumption. Virtually no reviewer actually recalculates the results of submitted manuscripts. Almost every reviewer judges results by their

plausibility and base their judgment on trust in the author's academic integrity. Results from prominent journals are often indiscriminately cited for years, or even decades, since critical replications are of no avail (seeing as serious shortcomings are rarely discovered). And if, indeed, results are replicated, it becomes apparent that publicized results often cannot be accurately replicated, especially in the social, behavioral and economic sciences. In most cases – as many tests have repeatedly shown – this is due to incomplete documentation or a complete lack of documentation of the underlying data or, at least, lack of publication thereof. Strictly speaking, these publications are unscientific. Further, even if a replication is successful, there is hardly a chance to publish the results – except for rare exceptions when spectacular cases of plagiarism are exposed.

The systematic misapprehension that lies in the objectification of academic achievements is that a formal quality test, the process of peer-reviewing, which is supposed to make selecting papers easier for journal editors, has contraindicated effects: it restrains consistent scientific quality, because the current mode of publication is designed to produce preferably new and innovative analyses. From the publishers' perspective, this is absolutely reasonable economically. The plain truth is: there is simply no room for tedious replications. Nevertheless, they are what researchers should be demanding.

Anyhow: cases of quasi-replication of research results can be observed in engineering and the applied natural sciences: replications of results are produced incidentally in everyday life: If a newly constructed bridge withstands, the efficiency of the new method, namely bridge building, is proven. Or the bridge collapses, and the efficiency of a new construction method could not be replicated.

But, this "replication by application" is seldom possible in basic research. In the social, behavioral, and economic sciences, replication studies do not have a tradition, because their results are often forecasts. These forecasts are necessarily imprecise, and their application rare. If a publicized result is not replicable, this is often accepted with a shrug. Then, it is insinuated that inadequate replicability is not due to bad analyses, but due to real-world changes that occurred meanwhile. Especially in the social sciences, everybody is busy blithely debating and postulating ever new hypotheses – in the arts and culture section of the newspapers, if need be –

instead of sincerely trying to uncover the causes for diverging empirical results and weak predictions.

Greater recognition of replication studies as distinct academic achievements, and greater chances for those to be published, would go towards significantly raising the bar for good scientific practice in empirical research. And this recognition of replication studies would have a further, side-effect that should not be underestimated: many interesting research data would become available as a by-product, which could be subsequently used for many other purposes. Secondary analysis of research data that employs new methods, or other assumptions and theories, can yield equally innovative results as the primary research that originally generated the data.

In order to achieve this, it is necessary to get all the relevant players on board:

- **Scientists and researchers**: As individuals, they do not have the power to shed the constraints of their careers. But, without the community of (individual) scientists, a better system will not be developed. Intense discussions in the science communities are necessary with regard to the different roles as authors and reviewers and other key players, especially research funders and publishers.

- The **Research funders**: Research funders should pay more attention to research data management in their application procedures. But: additional money going towards data provision and documentation should not be allocated for advanced or already completed projects. Instead, already during the planning phase, the funders should make sure that every project has a data management plan, which cogently outlines how to deal with the project's research data, how to make it (freely) available, the mode of documentation, and, last but not least, how this will be achieved as a part of the funded project. Furthermore, non-compliance with data management plans has to be sanctioned. In the last resort, researchers might have to be excluded from further funding until his or her data has been archived and made available.

- The **publishers** (and the **reviewers**): If tangible incentives are required to induce scientifically sound procedures, they have to pay out in the currency of our community: reputation gained through publications. One could ask provocatively, whether peer-reviewed journals that do not include availability of data for replication studies in their policies are scientific at all. At the moment, there is only a small number of journals that do exactly that and make sure the data are archived and made available. But it is still virtually impossible to publish the documentation of a data set or a replication study in *Science* or other triple-A journals, if it doesn't yield spectacular results. As a researcher, this is where you want your work published to advance your career. So, even a prestigious journal like *Science* should ask itself, whether a special section for replication studies could be established. This should not be done online, because that hardly builds reputation. It should be done in print!

3 Conclusions

Cooperation between researchers, research funders and peer-reviewed journals is a key approach – otherwise we will continue to hear merry-go-round rationals why science is still unscientific in the 21st century and nobody's to blame: the researchers as individuals cannot bring about change. They are subject to the incentive systems of their disciplines, which are, in turn, intertwined with the economic interests of the respective publishers. This should not go unchallenged by the academic elite.

Appendix

http://www.sciencemag.org/site/special/data-rep/
(downloaded on January, 9th 2012)

TAKE THE POLL

Ideally, scientists would fully disclose their own raw data and methods and also spend time replicating others' work. What would best ensure this good behavior? (Poll Closed)

- Recognition and rewards from institutions 6.07%
- Funding earmarked for replication studies 19.1%
- More publication by journals of data that confirm or refute previous work 44.86%
- Rewards from funders on subsequent grant applications for depositing sufficient details for replication (or penalties for noncompliance) 24.49%
- Other (add your ideas in the comments section) 5.48%