Software updates: the "unknown unknown" of the replication crisis

The replication crisis is largely concerned with known problems, such as the lack of replication standards, non-availability of data, or p-hacking. One hitherto unknown problem is the potential for software companies' changes to the algorithms used for calculations to cause discrepancies between two sets of reported results. **Anastasia Ershova** *and* **Gerald Schneider** *encountered this very problem in the course of their own replication test, and argue that software developers should take more responsibility for their role in the strengthening of replication standards.*

Speaking in 2002 about weapons of mass destruction, United States Secretary of Defense Donald Rumsfeld infamously distinguished between the "known unknowns" and the "unknown unknowns". The replication crisis that continues to engulf the social sciences is largely concerned with the known problems, so far including the lack of replication standards, the non-availability of data, p-hacking, and similar ills of an ever-growing science industry.

Admittedly, many of us have been aware of these problems for many years. Our field of study, political science, has been at the forefront of the replication movement that will hopefully dis-courage such behaviour in the long term. However, in the course of our work as editors of *European Union Politics*, we have discovered a problem that potentially undermines the reliability of many published studies and the credibility of those public policies that draw on these findings.

By trying to replicate the results of a conditionally accepted article, we uncovered discrepancies between the reported results calculated by the author and the ones obtained by us. These divergences spurred an intensive exchange between the author and us and, finally, resulted in the discovery that they are due to changes in an algorithm used by the (commercial) software company for calculations done with a certain estimator. The software company, which pressures universities and research institutes to buy the expensive updates of their statistical package every second year at least, reports that it has since modified its algorithm. Yet, the company does not justify which version of the program is the correct one to use in order to get as close as possible to the underlying true relationship. It could be the case that the new algorithm saves us computing times, while the older versions calculate more accurate coefficients.

We believe, based on this experience, that software developers should also play a role in the replication movement. Inconsistencies that are due to the selection of a faulty algorithm can, in the extreme, harm our lives. Just imagine a health intervention made based on a finding reached only due to the usage of an inappropriate algorithm. It is our opinion that the software company should receive the public blame for bad policymaking and ultimately be liable for damages it has induced. Software companies should also be forced to use the extra income generated by their frequent program updates to create a more encompassing documentation on the quality of their new and old products. Furthermore, perhaps before releasing a new version of the software for a broader usage, these companies should ensure it is bug-free by pre-testing it and thus guaranteeing the correctness of the produced estimations.

Yet, this new dimension in the replication debate should also lead to a further strengthening of replication standards. Researchers need to report which version of the software they used and, if this information is available, precisely when they last updated their software. In addition, they should be encouraged to replicate their findings with another software in the case that they are using a relatively newly developed estimator. A particular problem emerges through the development of estimators that are not yet official parts of a software package. Such freeware should only be used once an article in which this new estimator is presented has been published in a respected methods journal.

The further strengthening of replication standards we advocate here does not come freely. Recalculating findings sometimes takes several working days, and the possible usage of different versions of the same package at least doubles the effort replication teams must make. The additional costs are, at the moment, almost exclusively borne by the journal editors and their teams without any cost-sharing by the publishing industry. This amplifies the problem identified by UK physicist, Adrian Sutton: "What other industry receives its raw materials from its customers, gets those same customers to carry out the quality control of those materials, and then sells the same materials back to the customers at a vastly inflated price?" If we take replication seriously, we need to make all parties equally responsible – authors, reviewers, and editors, as well as the software developers and publishers.

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy if you have any concerns on posting a comment below.

Featured image credit: illustrade, via Pixabay (licensed under a CC0 1.0 license).

About the authors

Anastasia Ershova is a doctoral candidate at the Graduate School of Decision Sciences of the University of Konstanz. She has been a managing editor of European Union Politics until January 2018.

Gerald Schneider is Professor of International Politics at the University of Konstanz and Executive Editor of European Union Politics.