COMMENTARY
Thoughts & Opinion
BioEssays
www.bioessays-journal.com

Replications Everywhere

Why the replication crisis might be less severe than it seems at first

Stephan Guttinger

The debate about the replication crisis in the experimental sciences is based on two key claims, namely 1) that researchers rarely replicate existing data and 2) that if they attempt to do so they more often than not fail.

These claims have led to some serious soul-searching within the scientific community, the majority of the debate focusing on two issues: a) how can researchers be encouraged to perform more replications and b) how can it be ensured that fewer irreproducible data are created in the first place?

These are not idle debates because they could have serious consequences for the way in which research is conducted and funded. Some proposals made in response to the crisis go as far as suggesting that "blue-sky" basic research should be severely restricted in favor of research that is directly tied to practical outcomes.[1]

Here, I do not want to contribute to ongoing science policy debates, but instead question the very foundation on which these debates are built. In particular, I will argue that claim 1) is wrong because there is more replication going on in the experimental sciences than usually assumed. These replications, however, are completely ignored in the current debate because the analytic framework used cannot account for them. This also has implications for claim 2), because these additional replications are normally successful, making it likely that current estimates of the failure rate for replications (ranging from 50% to 80 %[2,3]) are too high. This suggests that there might be less of a crisis than some analysts claim.

This of course does not mean that issues such as better quality controls or the reporting of methods do not have to be addressed. Clearly, there is a lot that can be further improved in the experimental sciences. But what it means is that the doomsday-picture of a profound crisis that drives current calls for reform becomes less convincing. It also means that more focus should be put on getting as complete a picture of the status quo as possible before moving to the reform stage. The current debate, I claim, is built on flawed foundations.

S. Guttinger
Egenis, Centre for the Study of the Life Sciences
University of Exeter,
Byrne House, St German's Road, Exeter, EX4 4PJ, UK
E-mail: s.guttinger@exeter.ac.uk

S. Guttinger
CPNSS
London School of Economics and Political Science,
Lakatos Building, Houghton Street, London, WC2A 2AE, UK
E-mail: s.m.guettinger@lse.ac.uk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/bies.201800055.

DOI: 10.1002/bies.201800055

Explaining Trust in the Absence of Replication

The starting point for my reflections here is a point that rarely receives attention in the debate about the replication crisis, namely the fact that for decades scientists have not really been worried about the lack of replications in their own line of work: even though most researchers are familiar with how tricky it can be to replicate the work of others (or their own work for that matter), they were more or less content with how things were going (at least in the life sciences, on which this article focuses). Of course, there were debates about the quality of certain assays (the Far-Western blot would be a good example) or materials being used. But there certainly was little or no talk of a fundamental crisis.

What is more, researchers seem to have a deep trust in science itself and the data on which they are building in their own work. A recent survey conducted by *Nature* has confirmed this positive sentiment, showing that researchers still largely trust the data they are using, even though most of them acknowledge that, at some point in their career, they had problems replicating existing work.[4]

In the conceptual framework used to assess the replication crisis it is not possible to make sense of this trust. According to this framework, real trust could only come about if replications were performed before existing data are used. But because researchers do not set up replication studies, they do not seem to have good reason to trust what they are building on. They seem to move ahead blindly, simply hoping for the best.

The favored explanation for this apparently risky behavior of researchers is that they are forced to act this way by the perverse incentive structures that are in place within the sciences: the dominant publish-or-perish culture leaves them with no other choice but to throw all caution and critical attitude overboard and to use whichever data help them to get the next big story published. They simply do not have the time and money to thoroughly check what they are working with (in a more negative reading, the implication seems to be that many scientists lack integrity and let their ambition get the better of them).

Here, I want to propose a more positive take on why scientists have been doing things the way they have. I do not dispute that there are immense pressures on scientists to publish (and to publish well). And I am not saying that there are not some researchers who buckle under this pressure. But what I want to claim here is that scientists are not simply acting negligently, putting blind trust in existing data in order to get more money and to advance their own careers. Rather, the reason why scientists trust the data they are using is because there is a whole level of replication to which the current debate is completely blind. This form of replication, which I will call "micro-replication," is built into everyday research practice. Because of this it has slipped under the radar of most analysts as current consensus postulates that replications are always *add-ons* to regular experimentation; they are something that has to be done on top of what researchers normally do. This flawed conceptual framework leads to a distorted analysis of the status quo.

Micro-Replications: An Overlooked Source of Trust

In what follows I take my lead from a comment Stuart Firestein makes in his book "Failure," where he claims that "experiments get replicated because people from other labs use the published results and the methods in their own experiments."[5]

Firestein does not expand on this claim and he does not explain what such replication-in-practice looks like. Here, I want to show that an underappreciated part of the experimental process – namely experimental controls – provide a form of replication that has so far been overlooked in the debate on the replication crisis, namely the above-mentioned "micro-replications" (MRs).

To give an example of how MRs work I will look at a study published in *PLOS Biology* by Wang et al., who demonstrate that the choline transporter-like 1 (CTL1) protein plays a role in auxin regulation in *Arabidopsis*.[6] I chose this paper not because it has some unique features but because of the opposite: the paper is an exemplar of a standard research report. Two general features are of particular interest here: 1) like almost all studies published in biology journals these days, this study has not been designed as a replication of earlier findings. And 2), like most other studies, it builds on existing data to then develop its own message.

The last point is crucial: Wang et al. are not the first ones to characterize the CTL1 protein in *Arabidopsis*. A study published in *Nature Communications* in 2014 by Dettmer et al. already reported a (different) function of CTL1, namely its involvement in sieve plate development.[7] Wang et al. perform their study against this backdrop of existing knowledge about CTL1. This matters because in order to proceed the authors first establish that their mutant behaves as it should (meaning: as it was reported by others before). Against this validation of their system they then generate further knowledge about the roles CTL1 plays in plant physiology.

To validate their system – and hence their findings – Wang et al. first perform a series of positive controls. One such control is to show that their mutant displays the same root development defect that was already reported by Dettmer et al. (compare Figure 1D in Ref. [6] and Figure 1C in Ref. [7]). By doing so they demonstrate that the mutant plants they are working with are able to give insights into the effects of CTL1 mutations on plant development; as with any positive control, the point here is to show that their experimental system works in principle.

Importantly, by performing these experimental controls, the researchers perform a small-scale replication (MR) of existing data: they reproduce an earlier finding of another group using their own materials and methods (Wang et al., for instance, studied mutant plants that were generated using a different method than that used by Dettmer et al.). This, I claim, is an instantiation of the "replication-in-practice" that Firestein alludes to.

What is crucial in the context of the debate on the replication crisis is that the authors did not label their work as a replication study. This is a general feature of this form of replication: MRs are not mere add-ons to regular experimentation. Rather, they happen as an integral part of everyday research (in this case the functional characterization of a plant protein). This makes them a very powerful but easily overlooked form of replication.

As the name implies, micro-replications are small-scale replications that only reproduce certain aspects of earlier work. But this does not mean that they lack power, especially given how prevalent they are: because MRs are linked to a standard element of everyday research practice (i.e., positive controls) they represent a potentially very large set of (successful) replications that pervades the literature in the experimental life sciences. It is also, I claim, part of what makes researchers trust in their own work and that of others.

A New Basis on Which to Assess Reproducibility

The current debate about the reproducibility crisis is completely blind to the additional level of micro-replications, as the analytic framework used only thinks of replications as add-ons and not as part of regular experimentation. This leads analysts to conclude that researchers simply do not perform any replications and seem blindly to trust existing data. Once we expand our picture of replications, however, we end up with a very different assessment of the status quo – one in which with a significantly higher number of small-scale replications is performed on a regular basis. This changes the very foundation on which the debate about the replication crisis is built.

All of this is not to say that the current experimental sciences do not have to be improved. Surely, there are studies out there that cannot be reproduced and there are cases in which researchers use flawed materials (e.g., particular antibodies or contaminated cell lines). The ongoing debate about improving reporting and quality control is therefore still relevant and important. And if we can get researchers to do even more replications then surely that is also a good thing. Some existing data might for instance not be of the type that are used as a positive control or starting point for a new study and will therefore not be covered by MRs. There is a lot more that we need to learn about how MRs work and how prevalent (and powerful) they are.

But the discussion here shows that there might be less need to fundamentally revise the experimental sciences than current talk of a "crisis" often implies. There is good reason to assume that (successful) replications are much more prevalent. This, then, is not a call for inaction but for a revision of the foundations on which action is taken. In our efforts to further improve the experimental sciences, we have to make sure that we proceed with as complete an understanding of the status quo as possible. Taking into account MRs, I claim, will have to be part of this process.

Acknowledgements

The research leading to this paper has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n° 324186. The affiliations were updated on 22 June 2018 after original online publication.

Received: March 20, 2018
Published online: May 9, 2018

[1] D. Sarewitz, *New Atlantis* **2016**, *49*, 4.
[2] C. Begley, L. Ellis, *Nature* **2012**, *483*, 531.
[3] N. Vasilevsky, M. Brush, H. Paddock, L. Ponting, S. Tripathy, G. LaRocca, M. Haendel, *PeerJ* **2013**, *1*, e148.
[4] M. Baker, *Nature* **2016**, *533*, 452.
[5] S. Firestein, *Failure: Why Science Is So Successful*, Oxford University Press, Oxford, UK **2015**, p. 151.
[6] Y. Wang, L. Yang, Y. Tang, R. Tang, Y. Jing, C. Zhang, *PLoS Biol.* **2017**, *15*, e2004310.
[7] J. Dettmer, R. Ursache, A. Campilho, S. Miyashima, I. Belevich, S. O'Regan, D. Mullendore, S. Yadav, C. Lanz, L. Beverina, A. Papagni, *Nat. Commun*. **2014**, *5*, 1.