This is a post-peer-review, pre-copyedit version of an article published in Learning and Behavior. The final authenticated version is available online at: <http://dx.doi.org/10.3758/s13420-020-00442-5>

It's not just the animals that are STRANGE

Benjamin G. Farrar^{1,2*} and Ljerka Ostojić²

¹Department of Psychology, University of Cambridge, Cambridge, UK

⁵Institute for Globally Distributed Open Research and Education (IGDORE)

*Corresponding author

Email: bgf22@cam.ac.uk

Address: Department of Psychology, Downing Street, Cambridge CB2 3EB

Phone: 01223 333550

Summary

A “STRANGE” framework has been proposed to mitigate the effects of sampling biases in animal experiments (Webster & Rutz, 2020, *Nature*, 582, 337–340). While sampling biases are likely a major cause of poor replicability and generalizability in animal behavior research, a more comprehensive and cautious approach is needed to effectively guide decision making in animal research.

Webster and Rutz recently proposed a STRANGE framework to help researchers navigate a sampling bias problem in animal research. As an animal’s likelihood of being in a study sample might depend on its STRANGEness - its Social background, Trappability, Rearing history, Acclimation, Natural changes in responsiveness, Genetic make-up and Experience - this can affect the replicability and generalizability of an experiment’s results. Webster and Rutz suggest that journals should request a discussion of a sample’s STRANGEness in each publication and for this purpose they provide a series of questions that authors could consult.

The aim of the STRANGE proposal is laudable, and it effectively highlights how sampling biases are a key threat to the reliability and validity of animal behavior research. However, we believe that the STRANGE framework is limited conceptually (points 1 and 2) and even risks perpetuating some of the very biases it seeks to address (points 3 and 4).

1. It’s not just the animals that are STRANGE

While attempting to address sampling biases in animal behavior research, the STRANGE proposal misses many of the sampling biases that affect animal behavior research. By focusing on the experimental unit, the STRANGE problems ignore how experiments also sample from populations of settings, treatments and measurements, too. Sampling biases across these levels also constrain the generalizability and reliability of results (Yarkoni, 2019), and they should be integrated into any frameworks for navigating sampling biases in animal research. This is especially important as experimenters and apparatuses can have large influences on animal behavior, yet often only a single instance of each is sampled in an experiment. There are many angles to view the sampling problems

from: replicability, generalizability, theory testing, experimental design, and pseudo-replication and mixed-effects statistical analysis, and the STRANGE proposal would benefit from incorporating these to be most effective.

2. The STRANGE proposal needs a sampling theory

One method to unite the sampling bias problems is to take a sampling approach to experimental design (Machery, 2020). This begins with identifying the research objectives. Not all animal research aims to generalize to the whole species, and much of our understanding about animal behavior has come from studies with very “strange” animals. That a lab rat is not representative of some larger population of mammals does not prohibit the utility of research on it – rather, the STRANGEness of a sample is relative to the claim or theory at hand, and could even change over time. Ideally, a theory or claim will have certain boundary conditions that it can be tested within, and an effective framework would explicitly consider this by asking the following questions:

- a) What is the claim or theory being tested?
- b) What are the boundary conditions of this claim or theory, i.e., what are the populations of experimental units, settings, treatments and measurements that a valid test of the claim could sample from?
- c) In the current study or research programme, how representative are the samples of experimental units, settings, treatments and measurements representative of the populations specified by the claim or theory?

At what stage researchers wish to consider these questions may differ both within and between research programmes. For some research, sampling biases are best discussed at the level of the individual study, but for other research, sampling biases may be more effectively discussed at the level of the research programme, which may be more appropriate for research with captive animals where sampling biases may be similar across many studies. Ultimately, sampling biases should be discussed in research proposals, and considered in cost-benefit analyses alongside ethical frameworks and regulations.

3. Can we objectively evaluate a systems STRANGEness?

A key part of Webster and Rutz' proposal (2020, p. 340) is an '*objective evaluation of a system's STRANGEness*' when a manuscript is submitted for publication. However, this perpetuates a myth about scientific research: that there is a view from nowhere (Andrews, 2020). Of course, there will be many clear sampling biases that the STRANGE proposal will help researchers to discover, but there will be many more biases that researchers genuinely disagree about, many that they simply do not think of, and many that the social and cultural history and current norms of the field obscure: which field sites to use, which questions to ask, which species to test and importantly how to assess the sample's STRANGEness. This is a sampling problem of its own, and points to a wider need to consider and encourage the diversity of animal behavior research.

4. How the STRANGE proposal could hinder progress

Prescriptive measures, if not carefully implemented, risk exacerbating deeper biases in the scientific method of animal behavior science. Discussions about a sample's STRANGEness will lead to limited scientific improvement if they are simply added to the long list of caveats within a manuscript. As Goodhart's law dictates: when a measure becomes a target, it ceases to become a good measure. Rather than giving researchers a STRANGE procedure that could fall foul of the same ritualization and mis-use as procedures like null hypothesis significance testing, we should focus on the ultimate factors leading researchers to perform STRANGE research, i) lack of education, ii) lack of resources, and, iii) lack of incentives.

Conclusions

The STRANGE proposal is an opportunity to catalyze discussions and investigations into sampling biases in animal research, however as a framework it risks becoming a weak plaster for deeper issues in our scientific process. Nevertheless, the STRANGE acronym offers a potent and much needed reminder to consider sampling biases when performing animal research, and a more developed sampling perspective on experimental design could be integrated with systematic reporting guidelines, such as ARRIVE (Animal Research: Reporting of In Vivo Experiments; du Sert et al., 2020). This could provide

an effective short-term guide for researchers and journals to navigate sampling biases in animal research, and it will be strongest if it encompasses sampling biases throughout all levels of experimental design relative to the individual research question at hand. Whether this process is best performed at the level of the individual study, or as a more general evaluation of a research programme will vary within and between fields. Ultimately, a long-term and field-wide evaluation of the causes and consequences of sampling biases is required. This will involve developing the education, funding structures and incentives needed to allow researchers to generate samples that can achieve their research goals and promote a more visible discussion of what these goals are, alongside ethical cost-benefit analyses. This process will be slow, but necessary, for the problems highlighted by the STRANGE framework to receive the attention they deserve.

References

Andrews, K. (2020). *How to Study Animal Minds* (1st ed.). Cambridge University Press.

<https://doi.org/10.1017/9781108616522>

Machery, E. (2020). What is a Replication? [Accepted Preprint] *Philosophy of Science*.

<https://doi.org/10.1086/709701>

Sert, N. P. du, Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., Browne, W. J., Clark, A., Cuthill, I. C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S. T., Howells, D. W., Karp, N. A., Lazic, S. E., Lidster, K., MacCallum, C. J., Macleod, M., ... Würbel, H. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLOS Biology*, 18, e3000410. <https://doi.org/10.1371/journal.pbio.3000410>

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature*, 582, 337–340.

<https://doi.org/10.1038/d41586-020-01751-5>

Yarkoni, T. (2019). *The Generalizability Crisis* [Preprint]. PsyArXiv.

<https://doi.org/10.31234/osf.io/jqw35>