



The Legislative, Ethical, and Conceptual Importance of Replicability in Farm Animal Welfare Science

Christian Nawroth^{1,*} and Lorenz Gyga²

¹ Leibniz Institute for Farm Animal Biology, Institute of Behavioural Physiology, Dummerstorf, Germany

² Humboldt-Universität zu Berlin, Faculty of Life Sciences, Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences, Animal Husbandry & Ethology, Berlin, Germany

* Corresponding author (Email: nawroth.christian@gmail.com)

Citation – Nawroth, C., & Gyga, L. (2021). The legislative, ethical, and conceptual importance of replicability in farm animal welfare science. *Animal Behavior and Cognition*, 8(2), 247-250.
<https://doi.org/10.26451/abc.08.02.11.2021>

Abstract – In this commentary, we discuss three replicability issues that are specifically relevant to research regarding farm animal welfare: (1) Legislative action, and its potential economic consequences, should derive from robust and replicable research to benefit animals kept in an industrial setting. (2) From an ethical perspective, the use of relatively few additional animals in replication studies in farm animal welfare science can be justified in comparison with the high numbers of animals of the same species that are routinely kept in millions for production purposes. (3) Conceptually, studies in farm animal welfare science can take advantage of heterogeneous farm settings that can serve as the population for sampling and, accordingly, follow a framework that improves external validity and thus increases replicability. In light of these issues, animal welfare scientists should more strongly embrace measures to increase replicability in their research.

Keywords – Farm animals, Heterogenization, Replicability, Welfare

A considerable amount of published results will turn out to be false (Ioannidis, 2005). This might be due to questionable research practices, such as *p*-hacking (changing statistical methods until the *p*-value is under the threshold for significance) and HARKing (Hypothesizing After Results are Known; Head et al., 2015, or due to the chance aspect of the *p*-value and the dichotomy of defining a result as a “true” effect once a certain threshold is reached (e.g., $p < .05$; Cumming, 2012; Wasserstein et al., 2019). Nonetheless, a failure to replicate effects is an integral process of science, where hypotheses are tested and re-tested and some are accordingly identified as reliable and robust, while others are not (Earp & Trafimov, 2015). To better understand the presence or absence of specific effects, it is not only key to replicate a wide range of experiments, but also to identify field-specific research characteristics to assess the importance, necessity, as well as the limitations of replications.

In research conducted on animals, we often encounter a trade-off between the need to replicate studies and the concerns about the welfare of the animals involved (Prescott & Lidster, 2017). This is particularly relevant when experiments involve sub-optimal housing conditions (e.g., when animals are kept alone or when housing conditions are barren) and/or harmful treatments (e.g., by inducing pain or fear). One main area addressed in the field of applied ethology which faces the same trade-off is the welfare of animals under human care with a focus on farm animals. In farm animal welfare science, three

issues are specifically relevant when considering replicability: the legislative desirability, the ethical acceptability, and the intrinsic heterogeneity of study sites. Here we argue that: (1) animal welfare legislation and its potential economic consequences on animal keepers such as farmers has a specifically high demand for solid replication, (2) the ethical aspect of using animals for experimentation should be considered in the context that animals of the same species are routinely kept in millions in an industrial setting, and (3) testing animals at different facilities (e.g., farms) may provide some conceptual problems but, in effect, mainly benefits replicability by providing heterogeneous experimental contexts.

The Legislative and Economic Aspects of Replicability in Farm Animal Welfare Science

Farm animal welfare scientists assess how farmed animals cope with their environment (Gyga & Hillmann, 2018), which results in scientific knowledge on what animals want and like (Dawkins, 2004). This knowledge is meant to be used to improve housing conditions of animals kept under human care. The final step is political, though, and can be viewed as a societal compromise between economic and ethical factors. The knowledge on effects of specific interventions/treatments has the potential to lead to new recommendations on animal husbandry that directly influence how animals are kept or managed. This knowledge can then be implemented in either new (compulsory) legislation or in voluntary product labels. In both cases, convincing policymakers and practitioners can be difficult due to potential substantial economic cost that occurs when husbandry systems need to be adapted. These expenses usually need to be covered by farmers, who need to balance a (potential) improvement of their animals' welfare and their own economic constraints. If findings from applied research ought to be adapted to legislation and thus implemented into industrial settings, we need to be able to assure policymakers that the detected effects of treatments are reliable, robust, and meaningful. To keep the number of false-negative findings low, farm animal welfare scientists should assure that their studies are high-powered (e.g., through sample size and design decisions: Lazic, 2018; Rouder & Haaf, 2018). Also, once set, there will be only limited and slow possibilities for further legislative change in the near future. The most important aspect here is likely to be the consequences from the animals' point of view: any effect of improving welfare should be as certain as possible, optimally substantiated by a meta-analysis, because otherwise animals might not profit from any implemented measures. However, under special circumstances (e.g., if there are deadlines for the implementation of interventions) this certainty could be adjusted downwards (precautionary principle) if an opportunity for such an implementation will otherwise be missed. Replicability of applied research findings is therefore central due to the slow nature of legislative consequences and their potential high cost for the animal keeper. Replicability is hence likely to be of more practical relevance in applied ethology (and animal welfare science) than in fundamental animal behavior research. A similar issue regarding replicability is faced by behavioral research in translational medicine due to the extremely high cost of developing novel medications.

The Ethical Aspect of Replicability in Farm Animal Welfare Science

As in most other areas of animal behavior research, experiments in farm animal welfare science may confront test subjects with sub-optimal housing and/or management conditions (regarding the animals' needs and motivations). But unlike other areas, animal welfare scientists' explicit aim is to consider how these housing conditions affect the welfare of the tested animals and how these conditions can be improved. Often, the (control) treatments reflect the (minimal) requirements of the current legislation or best practice. Frequently, this means that these (control) treatments imply welfare conditions that are known already to be sub-optimal or minimal, at best. We can therefore ask whether it is ethical to conduct replications that expose farm animals to a minimal welfare standard. Additionally, it has to be taken into account in farm animal welfare science that millions (and depending on species even billions) of farm animals are exposed contemporaneously to the same minimal conditions in farming practice (FAO, 2017). From a utilitarian perspective, replicating experiments that have shown promising results regarding specific welfare parameters should be of high priority to assess whether these effects are

robust and meaningful. In order to reduce the number of animals in an experiment, the smallest effect size of interest (i.e., the point at which welfare is increased to a relevant degree) can be, in principle, defined in animal welfare science (although this may not be that easy in practice). In these cases (and if random variation is known), the minimum sample size to detect this effect size can be chosen in accordance with an adequate power analysis. Thus, the exposure of a relatively limited number of subjects to sub-optimal welfare conditions in replicating studies can lead to potential welfare improvements in a high number of farm animals once legislation has acknowledged the benefits of this very effect (see point 1).

The Conceptual Aspect of Replicability in Farm Animal Welfare Science

Experimental settings in farm animal welfare science often do not allow for identical direct replications because research is often conducted on farms, which provide a less-controlled environment than studies in laboratories. In addition, resources of institutions to create identical experimental laboratory settings are limited because large and expensive facilities are needed for the study of the relatively large farm animals. Although identical direct replications might thus not be possible, researchers can still aim to use closely related procedures and sample from similar enough populations of animals in similar settings to assess the reliability of the original experiment (Machery, 2019). But is noisiness in a system a problem - or rather a potential benefit? Farm animal welfare studies cover a broad range of study types (as do studies in other fields), ranging from highly experimental studies on a single farm (potentially with a single group of animals) to studies conducted on multiple farms, either in an epidemiological approach (e.g., comparing lame and non-lame cows in a quasi-natural experiment) or repeating a (simple) experiment on several farms. In the highly experimental studies on single groups, the high level of noise in farm settings may be a problem because idiosyncrasies of a specific farm may mask a potential effect. Furthermore, variables that are distinct to specific sites may lead to false positive findings, or findings that will not generalize beyond the original study. In contrast, multi-farm studies are quite common in applied ethology because it is feasible to recruit several (production) farms and statistical models have become more easily applicable in the last two decades for dealing with such hierarchically nested datasets (e.g., from Pinheiro & Bates, 2000, to Bürkner, 2017). Conducting the same type of observation or experiment on several farms may increase external validity, similar to multi-lab pre-clinical studies (see Voelkl et al., 2018, 2020). However, a multi-lab study with its own team of researchers at each lab may provide for increased external validity in comparison to a single team working on several farms (i.e., due to a lack of heterogeneity in experimenters conducting experiments, including their potential biases). Therefore, the latter approach may be more closely comparable to the approach of “systematic heterogenization of study samples” which is the idea that detected effects found in heterogenous contexts are more externally valid than those found in a more homogenous testing context and that those heterogeneous contexts can be actively produced by the researcher. This has been suggested as an easy step towards improved replication and at the same time applicable in a single lab by Voelkl et al. (2018, 2020). This heterogenization created by a multi-farm approach has the potential to drastically improve the replicability of detected effects (Voelkl et al., 2018) and improve generalizability (Yarkoni, 2019) and the severity of theoretical tests (Baribault et al., 2018; Mayo, 2018).

Conclusions

Due to sampling variance and the risk of false-positive and false-negative decisions, replications of studies are an essential part of scientific progress. We outline here that replicability has a special standing for research on farm animal welfare because of the practical legislative consequences based on the scientific claims. From a utilitarian perspective, the use of additional animals in replication studies in farm animal welfare science can be additionally justified because it means exposing only relatively few animals to conditions to which millions are exposed in an industrial setting, which all might profit from the research conducted. Finally, we highlight that studies in farm animal welfare science can easily follow a conceptual framework that improves external validity and thus increases replicability. In light of the

importance to base welfare decisions on secure findings and the fact that a high number of animals can profit from these findings, animal welfare scientists should strive to replicate their findings or, at least, implement the heterogenization of study populations to increase the replicability of the investigated effects.

Acknowledgments

We thank Anja Eggert, Edna Hillmann, Camille Raoult, and Claudio Tennie for comments on a previous version of the draft. We are also thankful to Benjamin Farrar and an anonymous reviewer for comments that helped to improve the manuscript. This work was supported by a grant from the Deutsche Forschungsgemeinschaft to CN (DFG, LA 1187/6-1).

References

- Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravec, Z., van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, *115*(11), 2607–2612. <https://doi.org/10.1073/pnas.1708285114>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Dawkins, M. S. (2004). Using behaviour to assess animal welfare. *Animal Welfare*, *13*, 3–7.
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*. <https://doi.org/10.3389/fpsyg.2015.00621>
- Food and Agriculture Organization (2017). *FAOSTAT livestock primary*. <http://www.fao.org/faostat/en> (accessed 12.18.19)
- Gygax, L., & Hillmann, E. (2018). “Naturalness” and its relation to animal welfare from an ethological perspective. *Agriculture*, *8*(9), 136. <https://doi.org/10.3390/agriculture8090136>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, *13*(3). <https://doi.org/10.1371/journal.pbio.1002106>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Lazic, S. E. (2018). Four simple ways to increase power without increasing the sample size. *Laboratory Animals*, *52*(6), 621–629. <https://doi.org/10.1177/0023677218767478>
- Machery, E. (2019). *What is a replication?* PhilSci Archive. <http://philsci-archive.pitt.edu/16553/>
- Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Cambridge University Press. <https://doi.org/10.1017/9781107286184>
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer. <https://doi.org/10.1007/b98882>
- Prescott, M. J., & Lidster, K. (2017). Improving quality of science through better animal welfare: The NC3Rs strategy. *Lab Animal*, *46*(4), 152–156. <https://doi.org/10.1038/lablan.1217>
- Rouder, J. N., & Haaf, J. M. (2018). Power, dominance, and constraint: A note on the appeal of different design traditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 19–26. <https://doi.org/10.1177/2515245917745058>
- Voelkl, B., Vogt, L., Sena, E. S., & Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biology*, *16*(2), e2003693. <https://doi.org/10.1371/journal.pbio.2003693>
- Voelkl, B., Altman, N. S., Forsman, A., Forstmeier, W., Gurevitch, J., Jaric, I., ... & Würbel, H. (2020). Reproducibility of animal research in light of biological variation. *Nature Reviews Neuroscience*, *21*(7), 384–393.
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$.” *The American Statistician*, *73*(sup1), 1–19. <https://doi.org/10.1080/00031305.2019.1583913>
- Yarkoni, T. (2020). *The generalizability crisis*. *The Behavioral and Brain Sciences*, 1–37. Advance online publication. <https://doi.org/10.1017/S0140525X20001685>