

## **The special case of replication studies in applied ethology**

Christian Nawroth<sup>1</sup>, Lorenz Gygax<sup>2</sup>

<sup>1</sup> Leibniz Institute for Farm Animal Biology, Institute of Behavioural Physiology, Dummerstorf, Germany

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

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 and HARKing (Head et al., 2012), 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.  $p < 0.05$ ; Cummings 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 falsified in due course. 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 animal behaviour research, 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). The field of applied ethology addresses the welfare of animals under human care with a focus on farm animals and faces the same trade-off. However, in comparison to more fundamental research, three issues are specifically relevant for this field: (1) animal welfare legislation and its potential economic consequences on animal keepers such as farmers, (2) the ethical aspect of using animals for experimentation that are routinely kept in millions in an industrial setting, and (3) the conceptual problems (and benefits) when testing animals at different facilities (e.g. farms).

### **1. The legislative and economic aspects of replicability in applied ethology**

Applied ethologists assess how farmed and captive animals cope with their environment (Gygax & 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 costs 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. Also, once set, there will be only limited and slow possibilities for further 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 do not profit from any implemented measures. 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 behaviour research. A similar issue regarding replicability is faced by behavioural research in translational medicine due to the extremely high cost of developing novel medications.

## **2. The ethical aspect of replicability in applied ethology**

As in most other areas of animal behaviour research, experiments in applied ethology may confront test subjects with sub-optimal housing and/or management conditions (regarding the animals' needs and motivations). But unlike other areas, applied ethologists' 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 applied ethology 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 the welfare parameters should be of high priority to assess whether these effects are robust and meaningful. 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).

### **3. The conceptual aspect of replicability in applied ethology**

Experimental settings in applied ethology often do not allow for 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. Hence, if replication studies are conducted in applied ethology, these reflect mostly conceptual rather than direct replications (Fidler & Wilcox, 2018). But is noisiness in a system a problem - or rather a potential benefit? Applied ethological 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, confounding variables that are distinct to specific sites may lead to false positive findings. 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). A multi-lab study with its own team of researchers at each lab may do so to a greater extent in comparison to a single team working on several farms. Therefore, the latter approach may be more closely comparable to the approach of “systematic heterogenization of study samples” that was suggested as an easy step towards improved replication and at the same time applicable in a single lab by Voelkl et al. (2018). This heterogenization created by a multi-farm approach can drastically improve the replicability of detected effects.

Due to the random nature of experiments and their limitations due to 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 applied ethological research because of the practical legislative consequences based on the scientific claims. From an utilitarian perspective, the use of additional animals in replication studies in applied ethology 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 applied ethology 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, applied ethologists should strive to replicate their findings or, at least, implement the heterogenization of study populations to increase 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. This work was supported by a grant from the Deutsche Forschungsgemeinschaft to CN (DFG, LA 1187/6-1).

## References

- Bürkner, P.-Ch. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1-28. doi: <https://doi.org/10.18637/jss.v080.i01>
- Cumming, G. (2012). *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge
- Dawkins, M. S. (2004). Using behaviour to assess animal welfare. *Animal welfare*, 13(1), 3-7. doi: <https://doi.org/10.18637/jss.v080.i01>
- FAO, 2017. FAOSTAT Livestock Primary. <http://www.fao.org/faostat/en> (accessed 12.18.19)
- Fidler, F. & Wilcox, J. (2018). Reproducibility of Scientific Results. *Stanford Encycl. Philos.* Winter 201. Metaphysics Research Lab, Stanford University.
- Gygax, L., & Hillmann, E. (2018). "Naturalness" and Its Relation to Animal Welfare from an Ethological Perspective. *Agriculture*, 8(9), 136. doi: <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). doi: <https://doi.org/10.1371/journal.pbio.1002106>
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi: <https://doi.org/10.1371/journal.pmed.0020124>

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York, NY. doi: <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. doi: <https://doi.org/10.1038/labam.1217>

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. doi: <https://doi.org/10.1371/journal.pbio.2003693>

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. doi: <https://doi.org/10.1080/00031305.2019.1583913>