



Automated Home-Cage Testing as a Tool to Improve Reproducibility of Behavioral Research?

Sophie Helene Richter*

Department of Behavioural Biology, University of Münster, Münster, Germany

Keywords: reproducibility crisis, behavioral data, automation, experimenter effect, systematic heterogenization

INTRODUCTION

Over the past few years, an intense discussion about the reproducibility of scientific findings in various life science disciplines arose (e.g., Baker, 2016; Scannell and Bosley, 2016; Voelkl and Würbel, 2016). Thereby, particular attention has been given to animal research, where irreproducibility prevalence rates have been estimated to range between 50 and 90% (Prinz et al., 2011; Collins and Tabak, 2014; Freedman et al., 2015). Highlighted under the umbrella term of the “reproducibility crisis,” mostly failures in the planning and conduct of animal experiments have been criticized to lead to invalid and hence irreproducible findings. To counteract these trends, animal scientists have repeatedly emphasized the need for a rethinking of current methodologies, mainly targeting aspects of the experimental design, but also of the reporting and publishing standards (see e.g., ARRIVE guidelines, Kilkenney et al., 2010; TOP guidelines, Nosek et al., 2016; du Sert et al., 2018; PREPARE guidelines, Smith et al., 2018).

However, as demonstrated by an already 20 years-old study, the use of thoroughly planned and well-reported protocols does not automatically lead to perfect reproducibility: In this study, three laboratories found remarkably different results when comparing the behavior of eight mouse strains in a battery of six conventionally carried out behavioral paradigms (Crabbe et al., 1999). It was hypothesized that “specific experimenters performing the testing were unique to each laboratory and could have influenced behavior of the mice. The experimenter in Edmonton, for example, was highly allergic to mice and performed all tests while wearing a respirator—a laboratory-specific (and uncontrolled) variable.” Following on from these thoughts, the importance of the experimenter as an uncontrollable background factor in the study of behavior moved into focus (e.g., Chesler et al., 2002a,b). Concomitantly, the use of automated test systems was promoted as a tool to reduce the experimenter’s influence and to improve the accuracy and reproducibility of behavioral data (e.g., Spruijt and DeVisser, 2006).

Against this background, the present opinion paper aims at (1) briefly discussing the role of the experimenter in animal studies, (2) investigating the advantages and disadvantages of automated test systems, (3) exploring the potential of automation for improving reproducibility, and (4) proposing an alternative strategy for systematically integrating the experimenter as a controlled variable in the experimental design. In particular, I will argue that systematic variation of personnel rather than rigorous homogenization of experimental conditions might benefit the external validity, and hence the reproducibility of behavioral data.

THE EXPERIMENTER AS AN UNCONTROLLABLE BACKGROUND FACTOR

As impressively illustrated by the above-mentioned multi-laboratory study, environmental conditions can exert a huge impact on behavioral traits. This has been particularly highlighted

OPEN ACCESS

Edited by:

Oliver Stiedl,
Vrije Universiteit
Amsterdam, Netherlands

Reviewed by:

Gernot Riedel,
University of Aberdeen,
United Kingdom

*Correspondence:

Sophie Helene Richter
richterh@uni-muenster.de

Specialty section:

This article was submitted to
Sleep and Circadian Rhythms,
a section of the journal
Frontiers in Neuroscience

Received: 02 March 2020

Accepted: 30 March 2020

Published: 24 April 2020

Citation:

Richter SH (2020) Automated
Home-Cage Testing as a Tool to
Improve Reproducibility of Behavioral
Research? *Front. Neurosci.* 14:383.
doi: 10.3389/fnins.2020.00383

in the context of behavioral genetics, where experimental factors have been observed to interact greatly with trait-relevant genes. In order to identify and rank potential sources of such variability, Chesler et al. initially applied a computational approach to a huge archival data set on baseline thermal nociceptive sensitivity in mice (Chesler et al., 2002a,b). This way, they systematically identified several experimental factors that affected nociception, including, for example, season, cage density, time of day, testing order, or sex. Most interestingly, however, this analysis revealed that a factor even more important than the mouse genotype (i.e., the treatment under investigation) was the experimenter performing the test. Following on from this initial finding, subsequent studies provided further empirical evidence for the influence of the experimenter on the outcome of behavioral tests (e.g., van Driel and Talling, 2005; Lewejohann et al., 2006; López-Aumatell et al., 2011; Bohlen et al., 2014). With the aim of further disentangling what exactly constitutes the experimenter effect, some of these studies concentrated on specific characteristics of the personnel working with the animals. In particular, they could show that certain characteristics, such as the sex of the experimenter (Sorge et al., 2014) or the animals' familiarity with the personnel (van Driel and Talling, 2005) may play a crucial role.

Moreover, with respect to behavioral observations and direct experimenter-dependent assessments (e.g., counting "head-dips" on the elevated plus maze), it cannot be ruled out that the human observer may evaluate observations inconsequently and that definitions of behavior in, for example, ethograms are interpreted in various ways. Training deficits as well as a lack of inter-observer reliability can thus be regarded as additional sources of the experimenter-induced variation (Spruijt and DeVisser, 2006; Bohlen et al., 2014). Furthermore, only short habituation periods may promote differential reactions of individual animals toward the observer. Irrespective of the precise features that account for the described experimenter effects, however, the research community has widely agreed upon the importance of this factor as an uncontrollable background factor in behavioral research.

THE USE OF AUTOMATED TEST SYSTEMS IN BEHAVIORAL STUDIES

To overcome this issue, voices became loud during the last years to increase the usage of automated and experimenter-free testing environments, particularly in behavioral studies. Thereby, automation is not only considered beneficial to reduce or even prevent the confounding impact of the "human element", but also to decrease the time-consuming efforts of human observers. Looking at the literature, two major research lines have been pursued to implement this idea further: (1) Automation of recording and test approaches, and (2) development of automated home cage phenotyping, or alternatively, test systems that are attached to home cages and can be entered on a voluntary basis. Whereas, the former involves testing the animal outside of the home cage and thus still requires handling by an experimenter (e.g., Horner et al., 2013), the latter enables a completely new route for monitoring behavior over long

periods of time within the familiar environment and without any need for human intervention (e.g., Jhuang et al., 2010). With regard to automated test systems used outside of the home cage, typical examples are touchscreen chambers, mainly used for the assessment of higher cognitive functions in rodents (e.g., Bussey et al., 2008, 2012; Krakenberg et al., 2019), Skinner boxes (e.g., Rygula et al., 2012), or more specifically targeted technologies, such as the automated maze task (Pioli et al., 2014), the automated open field test (Leroy et al., 2009), or the automated social approach task (Yang et al., 2011). Likewise, systems, such as the IntelliCage (e.g., Vannoni et al., 2014), the PhenoCube (e.g., Balci et al., 2013), or the PhenoTyper (e.g., De Visser et al., 2006) have been developed to track the behavior within the familiar home cage. Potential advantages of such automated compared to manual assessments include the continuous monitoring, particularly during the dark phase when mice are most active, the observation in a familiar and thus less stressful environment, and the examination of combinations of behaviors rather than single behaviors (Steele et al., 2007). The latter point has been particularly highlighted as being crucial for the behavioral characterization of rodent disease models, as signs of ill health, pain, and distress tend to be very subtle in these animals (Weary et al., 2009). Furthermore, letting animals self-pace their task progression from a home-cage has been shown to speed up learning and to increase test efficiency in complex tasks (e.g., 5-choice serial reaction time task, Rummelink et al., 2017). Lastly, the use of automated technologies allows animals to maintain some control over which resources they would like to interact with, a key advantage in terms of animal welfare (Spruijt and DeVisser, 2006). At the same time, automation may come with certain challenges: For example, many automated test systems are not yet adapted to group housing and thus may require single housing of the study subjects, at least during the observation phases. This in turn may critically impair the welfare of these individuals, and undermines the goal of refining housing conditions for social animals according to their needs (Richardson, 2012; but see also Bains et al., 2018). Likewise, even the best automation does not prevent the individual from being handled for animal care reasons, probably potentiating the stress experienced during these rare events. Although the increasing implementation of automated systems in behavioral studies thus brings about a number of advantages, there is still room and need for further improvement.

AUTOMATION AND REPRODUCIBILITY OF BEHAVIORAL DATA

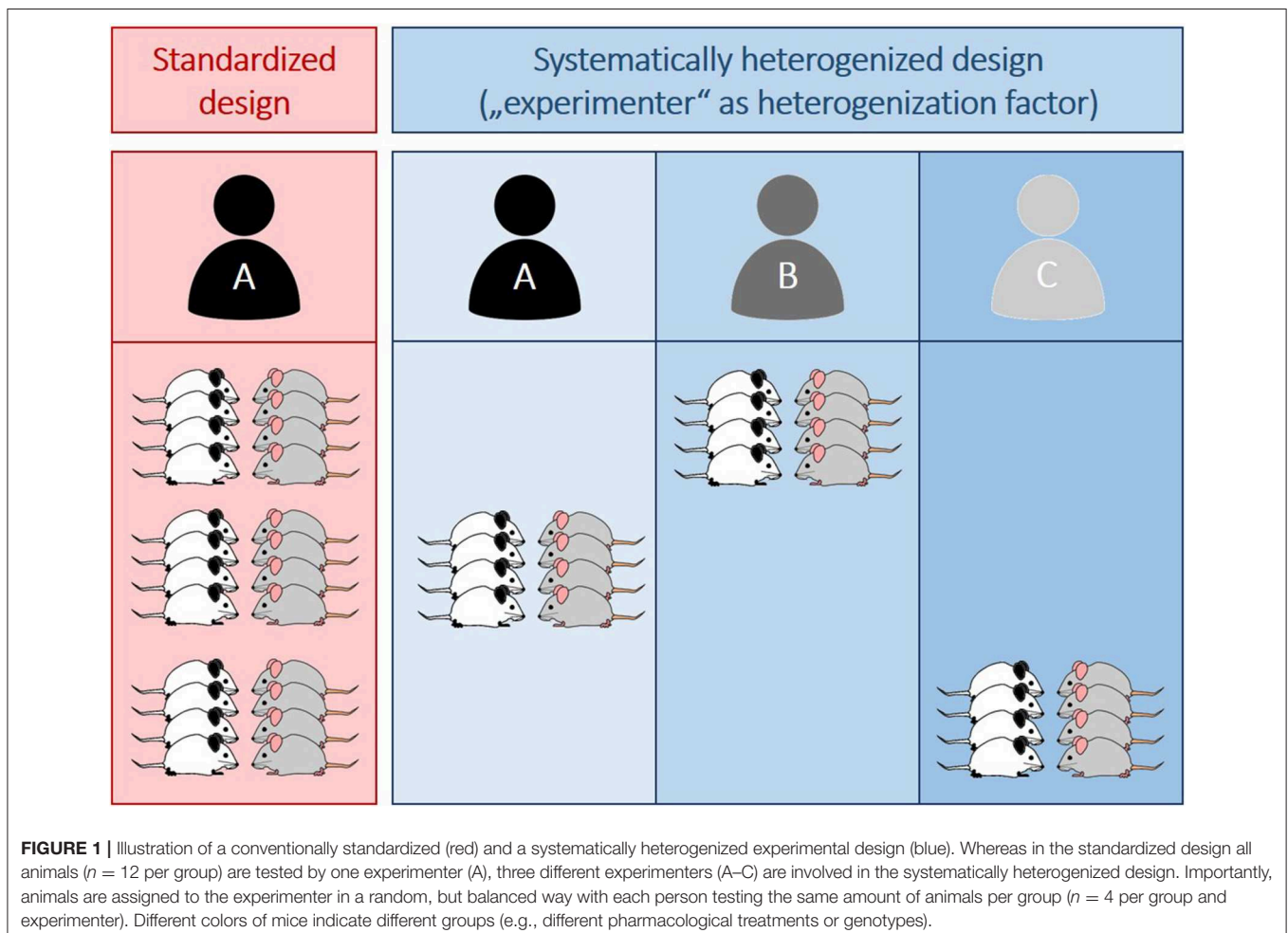
In light of the hotly discussed reproducibility crisis, automation is especially promoted as one potential way out of the problem. In particular, it has been argued that a computer algorithm, once programmed, and trained is consistent and unbiased and may thus reduce unwanted variation and hence contribute to improved comparability and reproducibility across studies and laboratories (Spruijt and DeVisser, 2006; Spruijt et al., 2014). In line with these arguments, a behavioral characterization of C57BL/6 and DBA/2 mice in the PhenoTyper indeed revealed

highly consistent strain differences in circadian rhythms across two laboratories (Robinson et al., 2018). Likewise, automated home-cage testing in IntelliCages was found to provide consistent behavioral and learning differences between three mouse strains across four laboratories, i.e., no significant laboratory-by-strain interactions could be detected (Krackow et al., 2010). Furthermore, comparing this system with conventional testing of mice in the open field and the water maze tests yielded more reliable results in the IntelliCages, even though the conventional tests were standardized strictly (Lipp et al., 2005). All of these studies indeed hint toward improved reproducibility through automation, suggesting that the absence of human interference during behavioral testing is a prominent advantage. However, systematic investigations on this topic are still scarce, in particular when it comes to comparisons to conventional approaches. Furthermore, significant behavioral variation has also been found to occur among genetically identical individuals that lived in the same “human-free” environment (Freund et al., 2013), indicating that the link between absence of human interference, reduced variation, and better reproducibility is not straightforward. As an alternative to improving reproducibility through automation, one may thus also think about turning

the experimenter effect into something “advantageous” by systematically considering this factor in the experimental design. So, what exactly is meant by this?

DISCUSSION—ALTERNATIVE STRATEGIES TO IMPROVE REPRODUCIBILITY

Following the above-presented logic, automated test systems reduce the influence of the experimenter, and may therefore be characterized by a higher degree of within-experiment standardization. Typically, it is argued that such increased standardization reduces variation, thereby improves the test sensitivity, and hence allows for detecting statistically significant effects with a lower number of animals (e.g., Richardson, 2012). At the same time, however, it has been pointed out that rigorous standardization limits the inference to the specific experimental conditions, thereby boosting any laboratory-specific deviations. Increasing the test sensitivity through rigorous standardization therefore comes at the cost of obtaining idiosyncratic results of limited external validity [referred to as “standardization fallacy” by Würbel (2000, 2002)]. Instead, the use of more heterogeneous



samples has been suggested to make study populations more representative and the results more “meaningful” (Richter et al., 2009, 2010, 2011; Voelkl and Würbel, 2016; Richter, 2017; Milcu et al., 2018; Voelkl et al., 2018; Bodden et al., 2019). According to this idea, the introduction of variation on a systematic and controlled basis (referred to as “systematic heterogenization” by e.g., Richter et al., 2009, 2010; Richter, 2017) predicts increased external validity and hence improved reproducibility.

As outlined above, increased automation may entail similar risks as it excludes one factor (i.e., the experimenter) known to induce or explain a lot of variation in behavioral studies. This way, conditions within experiments are more stringently homogenized, increasing the risk for obtaining spurious findings. Instead of trying to eliminate this variation, one may therefore think about systematically including it to improve the overall robustness of the data. Building on previous heterogenization studies (e.g., Richter et al., 2010; Bodden et al., 2019), this would simply mean to vary the within-experiment conditions by systematically involving more than just one experimenter. More precisely, instead of including one experimenter, who is responsible for testing all animals of one experiment (“conventional standardized design”, **Figure 1**), animals could for example be split in three equal groups (balanced for treatment, see also Bohlen et al., 2014), each tested by a different person (“systematically heterogenized design”, **Figure 1**).

From a practical perspective, such an approach may be associated with certain challenges, especially for small research groups with limited resources. For bigger research organizations or large-scale testing units, however, the organizational efforts

might increase only marginally. Balancing overall costs and benefits, such an experimenter-heterogenization may still represent an effective and easy-to-handle way to maximize the informative value of each single experiment (see Richter, 2017). Thus, rather than eliminating the uncontrollable factor “experimenter”, it could be turned into a controllable one that—systematically considered—may in fact benefit the outcome of behavioral studies.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

Research about systematic heterogenization and reproducibility is supported by a grant from the German Research Foundation (DFG) to SR (RI 2488/3-1). Furthermore, I would like to thank the Ministry of Innovation, Science and Research of the state of North Rhine-Westphalia (MIWF) for supporting the implementation of a professorship for behavioural biology and animal welfare (Project: Refinement of Animal Experiments).

ACKNOWLEDGMENTS

I would like to thank Vanessa von Kortzfleisch for her helpful comments on an earlier draft of this manuscript.

REFERENCES

- Bains, R. S., Wells, S., Sillito, R. R., Armstrong, J. D., Banks, G., and Nolan, P. M. (2018). Assessing mouse behaviour throughout the light/dark cycle using automated in-cage analysis tools. *J. Neurosci. Methods* 300, 37–47. doi: 10.1016/j.jneumeth.2017.04.014
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454. doi: 10.1038/533452a
- Balci, F., Oakeshott, S., Shamy, J. L., El-Khodori, B. F., Filippov, I., Mushlin, R., et al. (2013). High-throughput automated phenotyping of two genetic mouse models of Huntington’s disease. *PLoS Currents* 5:ENEURO.0141-17.2017. doi: 10.1371/currents.hd.124aa0d16753f88215776fba102ceb29
- Bodden, C., von Kortzfleisch, V. T., Karwinkel, F., Kaiser, S., Sachser, N., and Richter, S. H. (2019). Heterogenising study samples across testing time improves reproducibility of behavioural data. *Sci. Rep.* 9, 1–9. doi: 10.1038/s41598-019-44705-2
- Bohlen, M., Hayes, E., R., Bohlen, B., Bailoo, J., D., Crabbe, J. C., and Wahlsten, D. (2014). Experimenter effects on behavioral test scores of eight inbred mouse strains under the influence of ethanol. *Behav. Brain Res.* 272, 46–54. doi: 10.1016/j.bbr.2014.06.017
- Bussey, T. J., Holmes, A., Lyon, L., Mar, A. C., McAllister, K., Nithianantharajah, J., et al. (2012). New translational assays for preclinical modelling of cognition in schizophrenia: the touchscreen testing method for mice and rats. *Neuropharmacology* 62, 1191–1203. doi: 10.1016/j.neuropharm.2011.04.011
- Bussey, T. J., Padain, T., L., Skillings, E., A., Winters, B., Morton, A. J., and Saksida, L. M. (2008). The touchscreen cognitive testing method for rodents: how to get the best out of your rat. *Learn. Memory* 15, 516–523. doi: 10.1101/lm.987808
- Chesler, E. J., Wilson, S., G., Lariviere, W., R., Rodriguez-Zas, S., and Mogil, J. S. (2002a). Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neurosci. Biobehav. Rev.* 26, 907–923. doi: 10.1016/S0149-7634(02)00103-3
- Chesler, E. J., Wilson, S., G., Lariviere, W., R., Rodriguez-Zas, S., and Mogil, J. S. (2002b). Influences of laboratory environment on behavior. *Nat. Neurosci.* 5, 1101–1102. doi: 10.1038/nn1102-1101
- Collins, F. S., and Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature* 505, 612–613. doi: 10.1038/505612a
- Crabbe, J. C., Wahlsten, D., and Dudek, B. C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science* 284, 1670–1672. doi: 10.1126/science.284.5420.1670
- De Visser, L. R., Van Den Bos, Kuurman, W. W., Kas, M. J., and Spruijt, B. M. (2006). Novel approach to the behavioural characterization of inbred mice: automated home cage observations. *Genes Brain Behav.* 5, 458–466. doi: 10.1111/j.1601-183X.2005.00181.x
- du Sert, N. P., Hurst, V., Ahluwalia, A., Alam, S., Altman, D. G., Avey, M. T., et al. (2018). Revision of the ARRIVE guidelines: rationale and scope. *BMJ Open Sci.* 2:e000002. doi: 10.1136/bmjopen-2018-000002
- Freedman, L. P., Cockburn, I. M., and Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biol.* 13:e1002165. doi: 10.1371/journal.pbio.1002165
- Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., et al. (2013). Emergence of individuality in genetically identical mice. *Science* 340, 756–759. doi: 10.1126/science.1235294
- Horner, A. E., Heath, C. J., Hvosllef-Eide, M., Kent, B. A., Kim, C. H., Nilsson, S. R., et al. (2013). The touchscreen operant platform for testing learning and memory in rats and mice. *Nat. Protocols* 8:1961. doi: 10.1038/nprot.2013.122
- Jhuang, H., Garrote, E., Yu, X., Khilnani, V., Poggio, T., Steele, A., et al. (2010). Automated home-cage behavioural phenotyping of mice. *Nat. Commun.* 1:68. doi: 10.1038/ncomms1064

- Kilkenny, C., Browne, W., Cuthill, I. C., Emerson, M., and Altman, D. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* 8:e1000412. doi: 10.1371/journal.pbio.1000412
- Krackow, S., Vannoni, E., Codita, A., Mohammed, A., Cirulli, F., Branchi, I., et al. (2010). Consistent behavioral phenotype differences between inbred mouse strains in the IntelliCage. *Genes Brain Behav.* 9, 722–731. doi: 10.1111/j.1601-183X.2010.00606.x
- Krakenberg, V., Woigk, I., Rodriguez, L. G., Kästner, N., Kaiser, S., Sachser, N., et al. (2019). Technology or ecology? New tools to assess cognitive judgement bias in mice. *Behav. Brain Res.* 362, 279–287. doi: 10.1016/j.bbr.2019.01.021
- Leroy, T., Silva, M., D'Hooge, R. J., Aerts, M., and Berckmans, D. (2009). Automated gait analysis in the open-field test for laboratory mice. *Behav. Res. Methods* 41, 148–153. doi: 10.3758/BRM.41.1.148
- Lewejohann, L., Reinhard, C., Schrewe, A., Brandewiede, J., Haemisch, A., Görtz, N., et al. (2006). Environmental bias? Effects of housing conditions, laboratory environment and experimenter on behavioral tests. *Genes Brain Behav.* 5, 64–72. doi: 10.1111/j.1601-183X.2005.00140.x
- Lipp, H., Litvin, O., Galsworthy, M., Vyssotski, D., Zinn, P., Rau, A., et al. (2005). “Automated behavioral analysis of mice using INTELLICAGE: inter-laboratory comparisons and validation with exploratory behavior and spatial learning,” in *Proceedings of Measuring Behavior 2005, Noldus Information Technology* (Wageningen), 66–69.
- López-Aumatell, R., Martínez-Membrives, E., Vicens-Costa, E., Cañete, T., Blázquez, G., Mont-Cardona, C., et al. (2011). Effects of environmental and physiological covariates on sex differences in unconditioned and conditioned anxiety and fear in a large sample of genetically heterogeneous (N/Nih-HS) rats. *Behav. Brain Funct.* 7:48. doi: 10.1186/1744-9081-7-48
- Milcu, A., Puga-Freitas, R., Ellison, A. M., Blouin, M., Scheu, S., Freschet, G. T., et al. (2018). Genotypic variability enhances the reproducibility of an ecological study. *Nat. Ecol. Evol.* 2, 279–287. doi: 10.1038/s41559-017-0434-x
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Bowman, S., Breckler, S., et al. (2016). *Transparency and Openness Promotion (TOP) Guidelines*. Available online at: <https://osf.io/vj54c/>
- Pioli, E. Y., Gaskill, B. N., Gilmour, G., Tricklebank, M. D., Dix, S. L., Bannerman, D., et al. (2014). An automated maze task for assessing hippocampus-sensitive memory in mice. *Behav. Brain Res.* 261, 249–257. doi: 10.1016/j.bbr.2013.12.009
- Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10, 712–712. doi: 10.1038/nrd3439-c1
- Rommelink, E., Chau, U., Smit, A. B., Verhage, M., and Loos, M. (2017). A one-week 5-choice serial reaction time task to measure impulsivity and attention in adult and adolescent mice. *Sci. Rep.* 7:42519. doi: 10.1038/srep42519
- Richardson, C. A. (2012). Automated homecage behavioural analysis and the implementation of the three Rs in research involving mice. *Perspect. Lab. Anim. Sci.* 40, 7–9. doi: 10.1177/026119291204000513
- Richter, S. H. (2017). Systematic heterogenization for better reproducibility in animal experimentation. *Lab. Animal* 46, 343–349. doi: 10.1038/labana.1330
- Richter, S. H., Garner, J. P., Auer, C., Kunert, J., and Würbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nat. Methods* 7, 167–168. doi: 10.1038/nmeth0310-167
- Richter, S. H., Garner, J. P., and Würbel, H. (2009). Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nat. Methods* 6, 257–261. doi: 10.1038/nmeth.1312
- Richter, S. H., Garner, J. P., Zipser, B., Lewejohann, L., Sachser, N., Touma, C., et al. (2011). Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study. *PLoS ONE* 6:e0016461. doi: 10.1371/journal.pone.0016461
- Robinson, L., Spruijt, B., and Riedel, G. (2018). Between and within laboratory reliability of mouse behaviour recorded in home-cage and open-field. *J. Neurosci. Methods* 300, 10–19. doi: 10.1016/j.jneumeth.2017.11.019
- Ryglu, R., Pluta, H., and Popik, P. (2012). Laughing rats are optimistic. *PLoS ONE* 7:e51959. doi: 10.1371/journal.pone.0051959
- Scannell, J. W., and Bosley, J. (2016). When quality beats quantity: decision theory, drug discovery, and the reproducibility crisis. *PLoS ONE* 11:e0147215. doi: 10.1371/journal.pone.0147215
- Smith, A. J., Clutton, R. E., Lilley, E., Hansen, K. E., and Brattelid, T. (2018). PREPARE: guidelines for planning animal research and testing. *Lab. Animals* 52, 135–141. doi: 10.1177/0023677217724823
- Sorge, R. E., Martin, L., J., Isbester, K., A., Sotocinal, S., Rosen, S., Tuttle, A., H., et al. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* 11, 629–632. doi: 10.1038/nmeth.2935
- Spruijt, B. M., and DeVisser, L. (2006). Advanced behavioural screening: automated home cage ethology. *Drug Discov. Today* 3, 231–237. doi: 10.1016/j.ddtec.2006.06.010
- Spruijt, B. M., Peters, S. M., de Heer, R. C., Pothuizen, H. H., and van der Harst, J. E. (2014). Reproducibility and relevance of future behavioral sciences should benefit from a cross fertilization of past recommendations and today's technology: back to the future. *J. Neurosci. Methods* 234, 2–12. doi: 10.1016/j.jneumeth.2014.03.001
- Steele, A. D., Jackson, W. S., King, O. D., and Lindquist, S. (2007). The power of automated high-resolution behavior analysis revealed by its application to mouse models of Huntington's and prion diseases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1983–1988. doi: 10.1073/pnas.0610779104
- van Driel, K. S., and Talling, J. C. (2005). Familiarity increases consistency in animal tests. *Behav. Brain Res.* 159, 243–245. doi: 10.1016/j.bbr.2004.11.005
- Vannoni, E., Voikar, V., Colacicco, G., Sánchez, M. A., Lipp, P., and Wolfer, D. P. (2014). Spontaneous behavior in the social homecage discriminates strains, lesions and mutations in mice. *J. Neurosci. Methods* 234, 26–37. doi: 10.1016/j.jneumeth.2014.04.026
- Voelkl, B., Vogt, L., Sena, E. S., and Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol.* 16:e2003693. doi: 10.1371/journal.pbio.2003693
- Voelkl, B., and Würbel, H. (2016). Reproducibility crisis: are we ignoring reaction norms? *Trends Pharmacol. Sci.* 37, 509–510. doi: 10.1016/j.tips.2016.05.003
- Weary, D., Huzzey, J., and Von Keyserlingk, M. (2009). Board-invited review: Using behavior to predict and identify ill health in animals. *J. Animal Sci.* 87, 770–777. doi: 10.2527/jas.2008-1297
- Würbel, H. (2000). Behaviour and the standardization fallacy. *Nat. Genet.* 26, 263–263. doi: 10.1038/81541
- Würbel, H. (2002). Behavioral phenotyping enhanced-beyond (environmental) standardization. *Genes Brain Behav.* 1, 3–8. doi: 10.1046/j.1601-1848.2001.00006.x
- Yang, M., Silverman, J. L., and Crawley, J. N. (2011). Automated three-chambered social approach task for mice. *Curr. Protocols Neurosci.* 56, 8–26. doi: 10.1002/0471142301.ns0826s56

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with the author SR.

Copyright © 2020 Richter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.