ABC 2021, 8(2):251-262
Animal Behavior and Cognition
DOI: https://doi.org/10.26451/abc.08.02.12.2021

Considering Generalizability: A Lesson from Auditory Enrichment Research on Zoo Animals

Nicola Khan [a,b,*], and Claudia A.F. Wascher[c]

[a] Department of Psychology and Neuroscience, University of St Andrews
[b] School of Animal, Rural, and Environmental Studies, Nottingham Trent University
[c] School of Life Sciences, Anglia Ruskin University

*Corresponding author (Email: nicola.khan@ntu.ac.uk)

Citation – Khan, N., & Wascher, C. A. F. (2021). Considering generalizability: A lesson from auditory enrichment research on zoo animals. *Animal Behavior and Cognition, 8*(*2*), 251-262. https://doi.org/10.26451/abc.08.02.12.2021

Abstract – Environmental enrichment, particularly auditory enrichment, has recently gained attention as a potential method for reducing stress and encouraging a more diverse behavior repertoire in captive animals. However, the effects of auditory enrichment on behavior are inconclusive between different studies, and interpretation of behavior has proven difficult. In this commentary, we discuss different factors, such as small sample sizes and diversity of social groups, which might contribute to contradicting results. We then discuss the value of replication studies in animal behavior research and provide a framework and practical guidelines for developing independent replications prior to publication.

Keywords – Applied animal behavior, Auditory enrichment, Captive settings, Environmental enrichment, Reproducibility

The replication of empirical research is widely acknowledged to be a cornerstone of the scientific process. Over the past decade, a number of controversies have spurred a plethora of articles and conversations on transparency, study design and questionable research practices, and the role of replication in psychological science and related fields (Aarts et al., 2015; Ioannidis, 2005; Wiggins & Chrisopherson, 2019). Usually, a primary study shows significant findings, and the expectation is that these can be replicated when tested with new data and in new laboratories. However, in reality, many findings have been replicated less often than expected, leading to questions regarding the validity of the original or replication study, and/or whether failure to replicate actually indicates that there is no true effect after all (Maxwell et al., 2015; Shrout & Rodgers, 2018). This ‘replication crisis’ led to a concerted effort by the Open Science Collaboration to replicate 100 psychology studies from three top-tier journals (Aarts et al., 2015). Each study was replicated once, and depending on the criterion used, only 36 to 47% of the original studies were successfully replicated. However, it should be noted that concerns have been raised regarding the validity of the results due to wide prediction intervals, error, bias, and differences between the original and new sample populations (Gilbert et al., 2016; Patil et al., 2016).

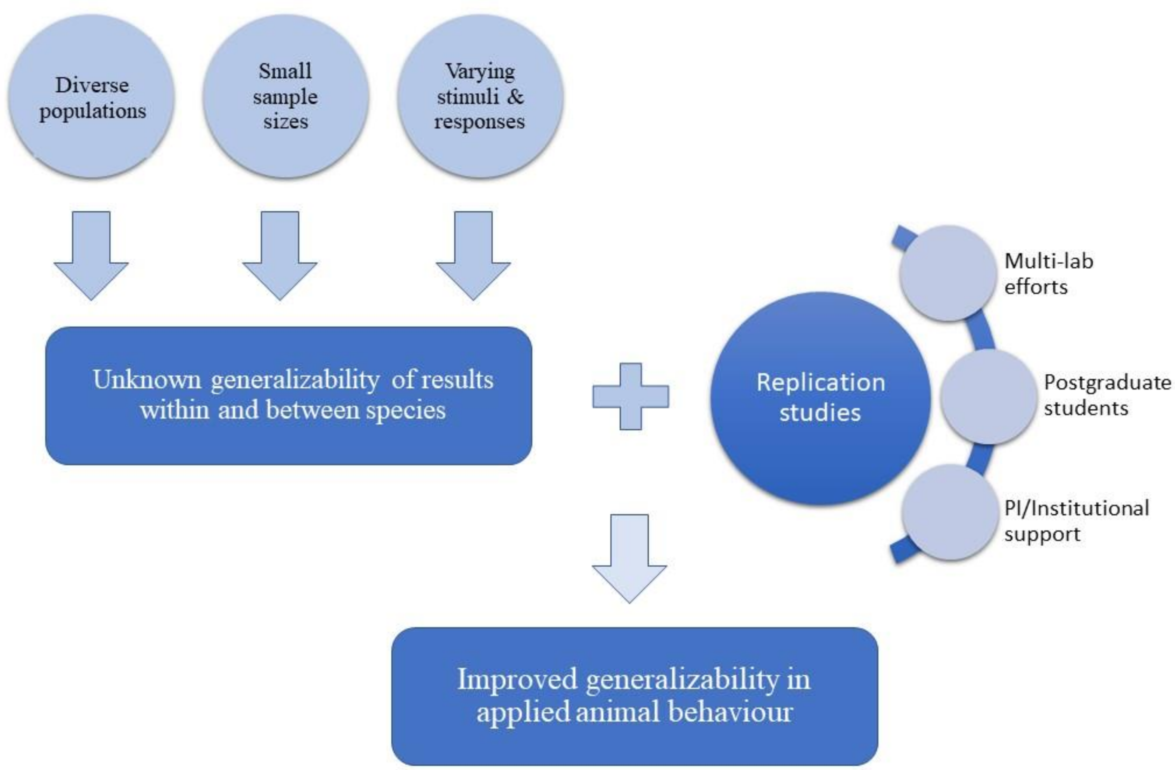
In comparison, the ‘Many Labs’ project took this a step further; 36 independent laboratories attempted to replicate each of 16 original psychology studies and pooled their data (Klein et al., 2014). This project resulted in much more optimistic figures; 85% of the original studies were successfully replicated (Gilbert et al., 2016; Klein et al., 2014). A later study by the same group successfully replicated 54% of 28 published findings (Klein et al., 2018). This indicates that ‘failures to replicate’ may not be failures at all but may simply be the result of single replication studies ultimately having too low a statistical power to identify true effects (Maxwell et al., 2015).

The replication crisis has forced other fields to reflect on the frequency and success with which they replicate previously published studies. Traditionally, the field of biology has a poor record regarding the publication of replication studies. A 2019 study reported that only 0.023% of studies in the general field of ecology and evolution were described by their authors as replications (Kelly, 2019), though this may be an artefact of promoting 'novelty' in publications. Similarly, researchers in the field of animal behavior or behavioral ecology seldom conduct replication studies; there is little culture of replications being encouraged or even expected (Kelly, 2006). Surprisingly, this includes the field of comparative cognition, a field closely related to psychology, where we would expect to see more replication studies (Farrar et al., 2020).

Here, we discuss the difficulty of conducting replications in applied animal behavior studies, using auditory enrichment studies in zoos as primary examples, and how encouraging replicates in this area could dramatically improve scientific progress and generalizability in the field of applied animal behavior. See Figure 1 for a representation of the value of replication research.

Figure 1

A Useful Model for the Value of Replication Research



Replication of Environmental Enrichment Studies

Millions of animals are housed in captive facilities worldwide, including farms, laboratories, zoos, and animal shelters. Unfortunately, captive conditions contain a wide array of potential stressors. For instance, the presence and noise from staff and visitors (Quadros et al., 2014) and/or social group size and species composition (in multi-species housing) may cause psychological stress (Price & Stoinski, 2007), spatial limits can restrict free movement and behavior (Polverino et al., 2015), and a lack of structural complexity, novelty, and/or repetitive diets may not provide sufficient mental or physical stimulation (Carlstead & Shepherdson, 2000; Morgan & Tromborg, 2007). As a result, animals may display stereotypy- fixed, repetitive behaviors that lack any obvious adaptive purpose (Powell et al., 2000; Singer, 2011). Fortunately, stereotypy can be prevented or reversed with appropriate environmental modification (Meehan et al., 2004). For example, forage-based enrichment, such as hiding or scattering food, can reduce or eliminate stereotypical behavior and promote behavioral diversity in a range of taxa (e.g., felids: Bashaw et al., 2003; Burgener et al., 2008; bears: Carlstead et al., 1991; pinnipeds: Fernandez & Timberlake, 2019; Hocking et al., 2015; primates: Baker, 1997; Bayne et al., 1991; mustelids: Malmkvist et al., 2013; aves: Meehan et al., 2003), and is now commonly used in captive facilities.

In recent years, auditory enrichment has been gaining attention as an additional form of environmental enrichment. Loud ambient noise, such as that found in animal facilities, has been shown to negatively impact a range of species (e.g., seahorses: Anderson et al., 2011, fish: Purser & Radford, 2011; Smith et al., 2004; Voellmy et al., 2014, birds: Blickley et al., 2012; Chloupek et al., 2009; Potvin & MacDougall-Shackleton, 2015, dogs: Gue et al., 1987, rats: Khasar et al., 2005, among others reviewed by Kight & Swaddle, 2011). As the value of music for the psychological well-being of humans is relatively well documented (e.g., MacDonald et al., 2012; McDermott et al., 2014; Whipple & Glynn, 1992), this has influenced the implementation of auditory enrichment for animals. A quarter of surveyed zoos provide auditory enrichment, such as animal vocalizations and music, for captive mammals at least once a week (Hoy et al., 2010) and radios are frequently played in animal facilities (Krohn et al., 2011).

Auditory provisioning may mask stressful sounds and improve animal welfare and behavioral diversity (Ogden et al., 1994). However, in animal facilities, there seems no clear consensus and standard practice regarding auditory enrichment protocols (e.g., stimuli vary considerably from ecologically relevant sounds to presentation of music or radio), all of which display varying sound statistics. Findings from auditory enrichment studies have been contradictory- likely due to a combination of small sample sizes and broad range of auditory stimuli. For example, auditory enrichment (rainforest sounds) resulted in reduced behavioral diversity in Senegal bush babies (*Galago senegalensis*) and two-toed sloths (*Choloepus didactylus*), while nine-banded armadillos (*Dasypus novemcinctus*) in the same study showed no behavioral changes (Clark & Melfi, 2012). Conflicting results occur even within species. For example, captive lowland gorillas (*Gorilla gorilla*) provided with ecologically relevant sounds have demonstrated reduced stereotypic stress-related behavior (Robbins & Margulis, 2014), no change in activity or anxiety-based behavior (Brooker, 2016), and increased locomotory behavior suggestive of agitation (Ogden et al., 1994) as well as a brief (<15 min) fear response when first presented with rainforest sounds (Wells et al., 2006). However, the same study later found that western lowland gorillas tended to engage in more relaxed behaviors and less stress-related behavior under the rainforest sound condition (Wells et al., 2006).

This variation in results of course raises the question, why do so many studies apparently contradict each other? First, sounds used in each of these studies differ in ‘ecological relevance’, and thus the sound statistics (and thus appeal) of the soundtracks would also vary. Further, many of these studies are characterized by small sample sizes with animals kept in different types of social groups. When we consider the lowland gorilla example discussed previously, each social group was completely different; a. four captive-born western lowland gorillas (two adult females, one adult male silverback, and one female infant, Robbins & Margulis, 2014), b. six lowland gorillas (one silverback, three adult females, one blackback male, one adolescent female. An infant male was also present, though was not behaviorally assessed in this study, Brooker, 2016), c. four lowland gorillas (adult male, adult female, and two infants, Ogden et al., 1994) and d. six (three males, three females; two wild-born, four captive-born) lowland gorillas aged between 8 and 41 years, Wells et al., 2006). There is little that can be done about this. Group sizes of certain species, for example, endangered species with long life cycles requiring high standards of housing, cannot simply be increased; thus, sample sizes are often out of control of individual researchers. Still, research on these species is very important, not only to ensure welfare of individuals in captivity, but also to gain general knowledge about the behavior of species, which for several reasons, is often difficult to study in the wild.

Individual responses are also likely to be context dependent. Differences in behavior between different studies may be a result of different housing conditions, and thus, differently enriched environments. For instance, solo-housed birds showed more stereotypy than group housed birds (Robbins & Margulis, 2016; Williams et al., 2017). This affects the opportunities for enrichment itself (e.g., whether a facility has the space or finances available required to provide highly enriched environments), as well as individual responses to enrichment as behavioral responses are influenced by early life history and general life experience. For example, early exposure to complex environments affects later development in deer mice (*Peromyscus maniculatus*: Hadley et al., 2006), in utero exposure to music in rats improves maze learning later in life (Rauscher et al., 1998), and environmental enrichment during rearing reduces fear levels and risk of injury in adult chickens (Reed et al., 1993). Rearing methods also play a crucial role; for instance, hand-reared birds show more stereotypic behaviors and interact less with enrichment than parent-reared birds (Williams et al., 2017), further emphasizing the need for considering context dependencies.

Furthermore, other studies test a very small sample of multiple species (often only one or two individuals per species) and describe substantial behavioral differences between species, which cannot necessarily be explained by socio-ecological factors (Robbins & Margulis, 2016; Williams et al., 2017). For example, Williams and colleagues (2017) presented ten zoo-housed psittacines of different species to six conditions of auditory stimulation (classical music, pop music, natural rainforest sounds, parrot sounds and a talking radio); unsurprisingly, behavioral responses were inconsistent. Similarly, the behavioral responses of three different African bird species to natural sounds, classical music, and rock music varied, with vocalizations and frequency of flying differing with music genre between species (Robbins & Margulis, 2016). Unfortunately, the small sample size here makes it difficult to interpret whether these forms of auditory enrichment impacted animal welfare. All of the behavioral effects described in these studies could be equally interpreted as signs of decreased welfare due to increased levels of stress (though it should be noted that none of the birds showed abnormal or negative stereotypies) or increased welfare by promoting more varied behavior.

This has resulted in a number of small-scale studies that attempt to gauge preference for different types of stimuli, most of which find high individual variation in preference. A classic study tested Java sparrows (*Lonchura oryzivora*; n=4) to determine whether they could distinguish between classical composers, and if they preferred the compositions by either J.S. Bach or A. Schoenberg (Watanabe & Nemoto, 1998). Two of the four showed a clear preference for Bach over Schoenberg, including when tested with novel pieces of music from the same composers. In comparison, the other two birds showed no preference. Unsurprisingly, other preference studies have also found considerable variation in preference across subjects (Fay & Miller, 2015; Fernandez et al., 2004; Mehrkam & Dorey, 2014), highlighting the need for additional research.

Within the field of applied animal behavior, and more specifically zoo-based studies, faithful replication of earlier studies is typically difficult because we study a diverse array of species in what can be highly variable settings. In comparison to other biological parameters, such as fecundity or longevity, behavior is naturally more difficult to quantify in an objective way. Indeed, when considering a study population in a zoo or research facility, given the small sample sizes, and variety of methods used for housing animals, as well as the potential impacts of developmental life experiences, it is impossible to reproduce exactly what other animal behaviorists have found. Such difficulties, however, do not diminish the importance of reproducibility. It is because of the specificity of these studies, *i.e*. this is what an *x*-year old male gorilla raised in *y* conditions does in response to *z*, that we need to carry out as many replications as possible in order to generalize behavior- particularly under captive conditions.

However, we also need to consider the ethical implications of running replicates, particularly in light of the guiding principles of the 3Rs: reduction, refinement, and replacement. Results of studies with small sample and effect sizes should not be expected to consistently replicate, especially as a likely publication bias towards positive results needs to be taken into consideration (Farrar et al., 2020).

Recently behavioral biologists have suggested the STRANGE framework to avoid sampling biases in animal behavior research (Webster & Rutz, 2020). The authors urge researchers to consider the social background, trappability and self-selection, rearing history, acclimation and habituation, natural changes in responsiveness, and genetic makeup of their study animals, as well as evaluating animal's life experiences when interpreting findings. Further, they argue that it is possible to mitigate STRANGE-related biases in studies by taking them into account when designing a study, declaring them in publications and discussing potential limitations (Webster & Rutz, 2020). We suggest that careful consideration and incorporation of the STRANGE framework when designing studies and interpreting results, together with replication studies, can potentially increase generalizability of applied animal behavior studies. In line with the 3Rs, each study conducted should use the minimum number of animals necessary without the loss of scientific rigor. This could be achieved by power analyses of pilot or previously published data (Cohen, 2013), which should be standard procedure at the stage of institutional ethical approval. Most examples we highlighted here consisted of zoo-based studies with small sample sizes and conflicting findings, emphasizing the need for more systematic attempts to replicate findings at larger scales. This will not only help to increase scientific knowledge about zoo-housed species, but also the general welfare of captive individuals.

Rethinking the Value of Replication Studies

While it is in everyone's interest that replication projects are conducted and the results publicly reported, there is no clear consensus about how to encourage these practices. In an ideal world, senior academics, research institutions, funding bodies, and journals would all acknowledge the importance of replication studies. There would be more focus on better understanding the generalizability and applicability of previous results in new or additional ecological systems. There is some merit to objections about allocating resources towards replication studies, which may be costly in terms of both financial and time costs, and away from alternative research. Furthermore, although natural science journal editors generally endorse replication studies as necessary, they are more difficult to publish (Madden et al., 1995), and, if published, tend to be published in journals with lower impact than the original publication (Frank & Saxe, 2012). As such, they are generally perceived as less valuable than original research, resulting in less recognition for the authors, and are less likely to be funded (Earp & Trafimow, 2015; Everett & Earp, 2015; Madden et al., 1995). This raises a conflict between self- and collective interest. Early career researchers (ECRs) are particularly affected by this social dilemma, as they are under strong selection pressure to publish high-impact work within the first couple of years of completing their Ph.D (Everett & Earp, 2015), which in the current academic environment is most likely achieved by publishing novel (ideally 'ground-breaking') studies. Failure to do so may result in failure to progress their careers and reduce the likelihood of future funding. Thus, ECRs that choose to invest time in replication studies are likely to reduce their employability in comparison to other ECRs that focus on novel experiments.

Other authors have suggested that replication studies should begin with undergraduate students, by introducing them to the replication crisis early on in their studies (Chopik et al., 2018), as well as involving undergraduate students in carrying out replications of new research as part of their coursework (Frank & Saxe, 2012; Grahe et al., 2012). Given that 100% of undergraduate psychology degrees in North America (from the 382 universities examined) offer courses in research methods, and 90% offer independent undergraduate research projects (Stoloff et al., 2010), the potential for contributing the results of replications towards large databases is substantial. The faculty that supervise these replication projects could report the number of projects and where these were submitted, thereby providing evidence of teaching-related research activity for their own evaluations (Grahe et al., 2012). Recognition of the value of replication projects within the field could raise the profile of such work and thus its value in meeting tenure requirements (Grahe et al., 2012).

Unfortunately, the type of replication studies that are suitable for undergraduate dissertation students will be, by necessity, those that are relatively simple and can be completed within a term.

Financially costly studies will be reproduced less often. Moreover, when considering the field of applied animal behavior, university and government-level restrictions, including the 3Rs, on working with animals will limit the type of studies that can be replicated. In comparison, Ph.D. students are more likely to have the skills (or be able to develop these with their supervisors) and additional time required to carry out high-quality and more complex replications (Everett & Earp, 2015). We believe that requiring Ph.D. students to carry out at least a single replication study in their field as part of their candidature would provide a multitude of benefits. Firstly, new Ph.D. students would be able to carry out and write up a research project early in their candidature; they would essentially be following an established methodology and (potentially also) statistical analysis, thereby increasing their confidence in conducting and critically evaluating research studies. As a result, students would gain in-depth knowledge of a study related to their upcoming original research. Indeed, their results may prove to be the foundation of their thesis and/or take their research in a new direction.

Moreover, working in a multi-lab group to carry out these collaborative replications would encourage early network development, which is crucial for ECRs. This approach will need the full support of their supervisors, with a clear structure of which study will be replicated as soon as the student starts their candidature, to allow time for ethics approval and necessary permits and training. However, hitting the ground running would allow for early write-up of their replication study as a thesis chapter- which could result in a relatively quick publication- all while still allowing them to carry out original research later in their Ph.D. For example, some of the published auditory enrichment studies in zoo animals could easily be replicated in other zoos that host the same species. Asian elephants (*Elephas maximus*), for instance, are a relatively common zoo animal, so there would be potential to run replicates of auditory enrichment studies (e.g., Wells & Irwin, 2008) and extend this topic further. Furthermore, as the results from these replication studies build up, patterns in generalizability will become clearer. From there, it is much more likely that we, as researchers, can push for funding bodies to consider additional funding streams- whether that is a push for independent direct replications, or support for schemes similar to the 'Many Labs' project- to support replication studies.

As a result, we feel that an integrated approach involving both undergraduate and postgraduate students would be most beneficial; this provides students with a rewarding learning environment that allows students to truly engage with research design and practice, while also providing large numbers of independent replication studies. If replications became a mainstream requirement of Ph.D. candidature, this minimizes the risk of some Ph.D. students being disadvantaged in comparison to others for conducting replications rather than focusing solely on original research (Everett & Earp, 2015). We acknowledge that this suggestion requires a cultural shift; journal editors, funding bodies, and academics will all need to come together to provide support for, and raise the value of, replication studies (Kelly, 2006; Nakagawa & Parker, 2015; Palmer, 2000).

Future Directions

Animal behavior and cognition studies in general often describe large individual differences in responses. These individual differences can present 'noise', *i.e*. confounding variables (for example, environmental factors such as housing or individual characteristics such as age or sex), affecting behavior rather than the factors of interest, such as environmental enrichment. On the other hand, individual differences in behavior can also present a meaningful response to a stimulus. Hence, as a first important step, future applied animal behavior research should focus on separating signals from noise; firstly, by identifying the behaviors of interest, and providing interpretations for any changes in behavior. A clear framework of which type of behavior acoustic enrichment is expected to elicit, such as relaxation versus excitation in the focal individuals, and how those responses are behaviorally quantified is crucial for a more consistent interpretation of behavioral responses. Secondly, confounding variables should be standardized as far as possible, or should be accounted for in statistical analysis. As such, within-subject designs in which individual responses to stimuli, for example auditory enrichment, are tested multiple times, provide a promising approach to produce replicable results (Farrar et al., 2020).

Despite the importance of generalization for effective implementation of environmental enrichment, few experiments have determined whether preferences generalize beyond the specific stimulus used for testing to a wider range of similar stimuli (e.g., Java sparrows generalize preferences for Bach; Watanabe & Nemoto 1998). Whereas it may be reasonable to assume that observed results generalize across individuals, populations, and contexts when there is no evidence to the contrary, the subtleties and complexities of context-dependent relationships within even a single species complicates this (Kelly, 2006; Klein et al., 2014). Failing to identify moderators and boundary conditions of effects may result in overgeneralizations across individuals, or when considering different environments, ecological contexts, or life history stages. For instance, individuals within a given population may behave differently throughout the year due to temporal changes in ecological and/or environmental variables (Foster & Endler, 1999). Cross-sectional studies that test enrichment preferences of animals of different ages combined with studies that test for change in preferences after repeated enrichment exposure within an age group could help tease apart these influences.

As discussed earlier, it is often too difficult to directly replicate or duplicate animal behavior studies when working with non-traditional organisms, particularly zoo-based animals. Studies are conducted on a wide variety of distantly related taxa and broad research questions, and research is handicapped by small sample sizes. In an ideal world, we would simply increase sample sizes when designing animal behavior studies or investigating enrichment practices. While this is, perhaps, feasible in captive studies that use traditional animal models, where large sample sizes can simply be ordered from suppliers, researchers that work on non-domesticated species and less traditional models are constrained by the availability of study animals, suitable housing, and technical support. Where studies are repeated within the field of ecology and evolution, researchers tend to use a different species or study system, resulting in partial replication, i.e., 'quasi-replication' (Palmer, 2000), or a more general conceptual replication (Schmidt, 2009). Conceptual replications are designed to assess the generalizability of results, usually by using similar (rather than identical) interventions, alternate measures of the outcome, or by studying populations that differ spatially, ecologically, or temporally (Fabrigar & Wegener, 2016). However, results from conceptual replications must be interpreted cautiously; if previous results are not replicated, it is not clear whether the original finding was spurious, or if the effect simply does not extend beyond the original experimental conditions (Fabrigar & Wegener, 2016; Shrout & Rodgers, 2018). Future studies could be designed to aid future replications and improve generalizability.

In other fields, *e.g.* psychology, biomedical sciences, where replication studies are more common, numerous laboratories and standardized conditions are available to replicate experiments under very similar conditions. For example, several studies confirm that the use of auditory enrichment increases neural plasticity in rats (Bose et al., 2010; Engineer et al., 2004; Kühlmann et al., 2018; Percaccio et al., 2007). Unfortunately, opportunities for such direct replications are often limited in behavior research that involves non-traditional model species. This issue could potentially be solved by collaborative efforts, as seen in other areas, such as the ManyPrimates projects, which was initiated to facilitate collaboration across study sites in primate cognition research (Altschul et al., 2019). Within initial ManyPrimates projects, researchers were able to include 176 individuals from 12 primate species housed at 11 sites across Africa, Asia, North America and Europe (Altschul et al., 2019). Similar ideas have been raised in other fields, where multiple labs have agreed to work together to test the reproducibility of their studies prior to publication (Schooler, 2014). Practical guidelines for such collaborative efforts already exist (Simons et al., 2014) and could be easily modified for environmental enrichment studies. In many areas of animal behavior and cognition, standardized experimental tests are already available, such as the detour task to assess motor-impulse control (Kabadayi et al., 2018), open field and novel object tests to assess animal personality (Carter et al., 2013; Perals et al., 2017), and string pulling to assess problem solving (Jacobs & Osvath, 2015). These test paradigms are applied in a comparable way in a wide variety of species and have been found to be repeatable in the same species.

In a collaborative effort, standardized experimental protocols could be conducted in zoos, research facilities, animal shelters or sanctuaries, and/or farms. For example, zebra finches (*Taeniopygia gutatta)* are housed in zoos and laboratories worldwide, thus replicates could be run across countries (to identify further across-population variation e.g., Griffith et al., 2017), or within regions, where variation in confounding factors are likely to be reduced, e.g., due to Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. This way, larger sample sizes are certainly possible and independent data from multiple institutions would provide replicates in themselves.

Conclusions

In our opinion piece we highlight challenges and importance of replication of behavioral studies conducted in zoos. We suggest that understanding the replication of studies be introduced to students at the undergraduate level, with Ph.D students encouraged to conduct replication studies related to their original research topics. We also suggest that labs work together to develop replication studies prior to publication and/or foster a work environment conducive to conducting replication studies and raising the value of replication studies beyond being simply confirmatory. We appreciate that this is a difficult task, particularly when studying non-traditional animals, but it is essential to ensure that current results in animal behavior are not simply artefacts of inadequate sample sizes and single study replications.

Acknowledgements

We would like to thank Dr. Christian Nawroth and three anonymous reviewers for their input, which helped strengthen this manuscript.

References

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., Duguid, S. J., Egelkamp, C. L., Fichtel, C., Fischer, J., Flessert, M., Hanus, D., Haun, D. B. M., Haux, L. M., Hernandez-Aguilar, R. A., Herrmann, E., Hopper, L. M., Joly, M., Kano, F., Keupp, S., … Watzek, J. (2019). Establishing an infrastructure for collaboration in primate cognition research. *PLOS ONE*, *14*(10), e0223675. https://doi.org/10.1371/journal.pone.0223675

Anderson, P. A., Berzins, I. K., Fogarty, F., Hamlin, H. J., & Guillette Jr, L. J. (2011). Sound, stress, and seahorses: the consequences of a noisy environment to animal health. *Aquaculture*, *311*(1–4), 129–138. https://doi.org/10.1016/J.AQUACULTURE.2010.11.013

Baker, K. C. (1997). Straw and forage material ameliorate abnormal behaviors in adult chimpanzees. *Zoo Biology*, *16*(3), 225–236. https://doi.org/10.1002/(sici)1098-2361(1997)16:3<225::aid-zoo3>3.0.co;2-c

Bashaw, M. J., Bloomsmith, M. A., Marr, M. J., & Maple, T. L. (2003). To hunt or not to hunt? A feeding enrichment experiment with captive large felids. *Zoo Biology*, *22*(2), 189–198. https://doi.org/10.1002/zoo.10065

Bayne, K., Mainzer, H., Dexter, S., Campbell, G., Yamada, F., & Suomi, S. (1991). The reduction of abnormal behaviors in individually housed rhesus monkeys (*Macaca mulatta*) with a foraging/grooming board. *American Journal of Primatology*, *23*(1), 23–35. https://doi.org/10.1002/ajp.1350230104

Blickley, J. L., Word, K. R., Krakauer, A. H., Phillips, J. L., Sells, S. N., Taff, C. C., Wingfield, J. C., & Patricelli, G. L. (2012). Experimental chronic noise is related to elevated fecal corticosteroid metabolites in lekking male greater sage-grouse (*Centrocercus urophasianus*). *PLOS ONE*, *7*(11), e50462. https://doi.org/10.1371/journal.pone.0050462

Bose, M., Muñoz-llancao, P., Roychowdhury, S., Nichols, J. A., Jakkamsetti, V., Porter, B., Byrapureddy, R., Salgado, H., Kilgard, M. P., Aboitiz, F., Dagnino-Subiabre, A., & Atzori, M. (2010). Effect of the environment on the dendritic morphology of the rat auditory cortex. *Synapse*, *64*(2), 97–110. https://doi.org/10.1002/syn.20710

Brooker, J. S. (2016). An investigation of the auditory perception of western lowland gorillas in an enrichment study. *Zoo Biology*, *35*(5), 398–408. https://doi.org/10.1002/zoo.21312

Burgener, N., Gusset, M., & Schmid, H. (2008). Frustrated appetitive foraging behavior, stereotypic pacing, and fecal glucocorticoid levels in snow leopards (*Uncia uncia*) in the Zurich Zoo. *Journal of Applied Animal Welfare Science*, *11*(1), 74–83. https://doi.org/10.1080/10888700701729254

Carlstead, K., Seidensticker, J., & Baldwin, R. (1991). Environmental enrichment for zoo bears. *Zoo Biology*, *10*(1), 3–16. https://doi.org/10.1002/zoo.1430100103

Carlstead, K., & Shepherdson, D. (2000). Alleviating stress in zoo animals with environmental enrichment. In G. P. Moberg & J. A. Mench (Eds.). *The biology of animal stress: Basic principles and implications for animal welfare* (pp. 337–354). CABI. https://doi.org/10.1079/9780851993591.0337

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioral ecologists measuring? *Biological Reviews*, *88*(2), 465–475. https://doi.org/10.1111/brv.12007

Chloupek, P., Voslářová, E., Chloupek, J., Bedáňová, I., Pištěková, V., & Večerek, V. (2009). Stress in broiler chickens due to acute noise exposure. *Acta Veterinaria Brno*, *78*(1), 93–98. https://doi.org/10.2754/avb200978010093

Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, *45*(2), 158–163. https://doi.org/10.1177/0098628318762900

Clark, F. E., & Melfi, V. A. (2012). Environmental enrichment for a mixed-species nocturnal mammal exhibit. *Zoo Biology*, *31*(4), 397–413. https://doi.org/10.1002/zoo.20380

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*, 621. https://doi.org/10.3389/fpsyg.2015.00621

Engineer, N. D., Percaccio, C. R., Pandya, P. K., Moucha, R., Rathbun, D. L., & Kilgard, M. P. (2004). Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons. *Journal of Neurophysiology*, *92*(1), 73–82. https://doi.org/10.1152/jn.00059.2004

Everett, J. A. C., & Earp, B. D. (2015). A tragedy of the (academic) commons: Interpreting the replication crisis in psychology as a social dilemma for early-career researchers. *Frontiers in Psychology*, *6*, 1152. https://doi.org/10.3389/fpsyg.2015.01152

Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*, *66*, 68–80. https://doi.org/10.1016/j.jesp.2015.07.009

Farrar, B. G., Boeckle, M., & Clayton, N. S. (2020). Replications in comparative cognition: what should we expect and how can we improve? *Animal Behavior and Cognition*, *7*(1), 1-22. https://doi.org/10.26451/abc.07.01.02.2020

Fay, C., & Miller, L. J. (2015). Utilizing scents as environmental enrichment: preference assessment and application with Rothschild giraffe. *Animal Behavior and Cognition*, *2*(3), 285–291. https://doi.org/10.12966/abc.08.07.2015

Fernandez, E. J., Dorey, N. R., & Rosales-Ruiz, J. (2004). A two-choice preference assessment with five cotton-top tamarins (*Saguinus oedipus*). *Journal of Applied Animal Welfare Science*, *7*(3), 163–169. https://doi.org/10.1207/s15327604jaws0703_2

Fernandez, E. J., & Timberlake, W. (2019). Foraging devices as enrichment in captive walruses (*Odobenus rosmarus*). *Behavioural Processes*, *168*, 103943. https://doi.org/10.1016/j.beproc.2019.103943

Foster, S. A., & Endler, J. A. (Eds.) (1999). *Geographic variation in behaviour: Perspectives on evolutionary mechanisms*. Oxford University Press.

Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, *7*(6), 600–604. https://doi.org/10.1177/1745691612460686

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science.” *Science*, 351(6277), 1037. https://doi.org/10.1126/science.aad7243

Grahe, J. E., Reifman, A., Hermann, A. D., Walker, M., Oleson, K. C., Nario-Redmond, M., & Wiebe, R. P. (2012). Harnessing the undiscovered resource of student research projects. *Perspectives on Psychological Science*, *7*(6), 605–607. https://doi.org/10.1177/1745691612459057

Griffith, S. C., Crino, O. L., Andrew, S. C., Nomano, F. Y., Adkins-Regan, E., Alonso-Alvarez, C., Bailey, I. E., Bittner, S. S., Bolton, P. E., Boner, W., Boogert, N., Boucaud, I. C. A., Briga, M., Buchanan, K. L., Caspers, B. A., Cichoń, M., Clayton, D. F., Derégnaucourt, S., Forstmeier, W., … Williams, T. D. (2017). Variation in reproductive success across captive populations: Methodological differences, potential biases and opportunities. *Ethology*, *123*(1), 1–29. https://doi.org/10.1111/eth.12576

Gue, M., Fioramonti, J., Frexinos, J., Alvinerie, M., & Bueno, L. (1987). Influence of acoustic stress by noise on gastrointestinal motility in dogs. *Digestive Diseases and Sciences*, *32*(12), 1411–1417. https://doi.org/0163-2116/87/1200-1411505.00/0

Hadley, C., Hadley, B., Ephraim, S., Yang, M., & Lewis, M. H. (2006). Spontaneous stereotypy and environmental enrichment in deer mice (*Peromyscus maniculatus*): Reversibility of experience. *Applied Animal Behaviour Science*, *97*(2–4), 312–322. https://doi.org/10.1016/j.applanim.2005.08.006

Hocking, D. P., Salverson, M., & Evans, A. R. (2015). Foraging-based enrichment promotes more varied behavior in captive Australian fur seals (*Arctocephalus pusillus doriferus*). *PLoS ONE*, *10*(5), e0124615. https://doi.org/10.1371/journal.pone.0124615

Hoy, J. M., Murray, P. J., & Tribe, A. (2010). Thirty years later: Enrichment practices for captive mammals. *Zoo Biology*, *29*(3), 303–316. https://doi.org/10.1002/zoo.20254

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

Jacobs, I. F., & Osvath, M. (2015). The string-pulling paradigm in comparative psychology. *Journal of Comparative Psychology*, *129*(2), 89–120. https://doi.org/10.1037/a0038746

Kabadayi, C., Bobrowicz, K., & Osvath, M. (2018). The detour paradigm in animal cognition. *Animal Cognition*, *21*, 21–35. https://doi.org/10.1007/s10071-017-1152-0

Kelly, C. D. (2006). Replicating empirical research in behavioral ecology: How and why it should be done but rarely ever is. *Quarterly Review of Biology*, 81(3), 221–236. https://doi.org/10.1086/506236

Kelly, C. D. (2019). Rate and success of study replication in ecology and evolution. *PeerJ*, *2019*(9), e7654. https://doi.org/10.7717/peerj.7654

Khasar, S. G., Green, P. G., & Levine, J. D. (2005). Repeated sound stress enhances inflammatory pain in the rat. *Pain*, *116*(1), 79–86. https://doi.org/https://doi.org/10.1016/j.pain.2005.03.040

Kight, C. R., & Swaddle, J. P. (2011). How and why environmental noise impacts animals: An integrative, mechanistic review. *Ecology Letters*, *14*(10), 1052–1061. https://doi.org/10.1111/j.1461-0248.2011.01664.x

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Krohn, T. C., Salling, B., & Hansen, A. K. (2011). How do rats respond to playing radio in the animal facility? *Laboratory Animals*, *45*(3), 141–144. https://doi.org/10.1258/la.2011.010067

Kühlmann, A. Y. R., de Rooij, A., Hunink, M. G. M., de Zeeuw, C. I., & Jeekel, J. (2018). Music affects rodents: A systematic review of experimental research. *Frontiers in Behavioral Neuroscience*, 12, 301. https://doi.org/10.3389/fnbeh.2018.00301

MacDonald, R., Kreutz, G., & Mitchell, L. (2012). *Music, health, and wellbeing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586974.001.0001

Madden, C. S., Easley, R. W., & Dunn, M. G. (1995). How journal editors view replication research. *Journal of Advertising*, *24*(4), 77–87. https://doi.org/10.1080/00913367.1995.10673490

Malmkvist, J., Palme, R., Svendsen, P. M., & Hansen, S. W. (2013). Additional foraging elements reduce abnormal behaviour - fur-chewing and stereotypic behaviour - in farmed mink (*Neovison vison*). *Applied Animal Behaviour Science*, *149*(1–4), 77–86. https://doi.org/10.1016/j.applanim.2013.10.001

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist,* 70 (6), 487–498. https://doi.org/10.1037/a0039400

McDermott, O., Orrell, M., & Ridder, H. M. (2014). The importance of music for people with dementia: The perspectives of people with dementia, family careers, staff and music therapists. *Aging & Mental Health*, *18*(6), 706–716. https://doi.org/10.1080/13607863.2013.875124

Meehan, C. L., Garner, J. P., & Mench, J. A. (2004). Environmental enrichment and development of cage stereotypy in Orange-winged Amazon parrots (*Amazona amazonica*). *Developmental Psychobiology*, *44*(4), 209–218. https://doi.org/10.1002/dev.20007

Meehan, C. L., Millam, J. R., & Mench, J. A. (2003). Foraging opportunity and increased physical complexity both prevent and reduce psychogenic feather picking by young Amazon parrots. *Applied Animal Behaviour Science*, *80*(1), 71–85. https://doi.org/10.1016/S0168-1591(02)00192-2

Mehrkam, L. R., & Dorey, N. R. (2014). Is preference a predictor of enrichment efficacy in Galapagos tortoises (*Chelonoidis nigra*)? *Zoo Biology*, *33*(4), 275–284. https://doi.org/10.1002/zoo.21151

Morgan, K. N., & Tromborg, C. T. (2007). Sources of stress in captivity. *Applied Animal Behaviour Science*, *102*(3–4), 262–302. https://doi.org/10.1016/j.applanim.2006.05.032

Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, *13*(1), 88. https://doi.org/10.1186/s12915-015-0196-3

Ogden, J. J., Lindburg, D. G., & Maple, T. L. (1994). A preliminary study of the effects of ecologically relevant sounds on the behaviour of captive lowland gorillas. *Applied Animal Behaviour Science*, *39*(2), 163–176. https://doi.org/10.1016/0168-1591(94)90136-8

Palmer, A. R. (2000). Quasi-replication and the contract of error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics*, *31*(1), 441–480. https://doi.org/10.1146/annurev.ecolsys.31.1.441

Patil, P., Peng, R. D., & Leek, J. T. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspectives on Psychological Science*, *11*(4), 539–544. https://doi.org/10.1177/1745691616646366

Perals, D., Griffin, A. S., Bartomeus, I., & Sol, D. (2017). Revisiting the open-field test: What does it really tell us about animal personality? *Animal Behaviour*, *123*, 69–79. https://doi.org/10.1016/j.anbehav.2016.10.006

Percaccio, C. R., Pruette, A. L., Mistry, S. T., Chen, Y. H., & Kilgard, M. P. (2007). Sensory experience determines enrichment-induced plasticity in rat auditory cortex. *Brain Research*, *1174*(1), 76–91. https://doi.org/10.1016/j.brainres.2007.07.062

Polverino, G., Manciocco, A., Vitale, A., & Alleva, E. (2015). Stereotypic behaviors in *Melopsittacus undulatus*: Behavioral consequences of social and spatial limitations. *Applied Animal Behaviour Science*, *165*, 143–155. https://doi.org/10.1016/j.applanim.2015.02.009

Potvin, D. A., & MacDougall-Shackleton, S. A. (2015). Traffic noise affects embryo mortality and nestling growth rates in captive zebra finches. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, *323*(10), 722-730. https://doi.org/10.1002/jez.1965

Powell, S. B., Newman, H. A., McDonald, T. A., Bugenhagen, P., & Lewis, M. H. (2000). Development of spontaneous stereotyped behavior in deer mice: Effects of early and late exposure to a more complex environment. *Developmental Psychobiology*, *37*(2), 100–108. https://doi.org/10.1002/1098-2302(200009)37:2<100::AID-DEV5>3.0.CO;2-6

Price, E. E., & Stoinski, T. S. (2007). Group size: Determinants in the wild and implications for the captive housing of wild mammals in zoos. *Applied Animal Behaviour Science*, *103*(3–4), 255–264. https://doi.org/10.1016/j.applanim.2006.05.021

Purser, J., & Radford, A. N. (2011). Acoustic noise induces attention shifts and reduces foraging performance in three-spined sticklebacks (*Gasterosteus aculeatus*). *PLoS One*, *6*(2), e17478. https://doi.org/10.1371/journal.pone.0017478

Quadros, S., Goulart, V. D. L., Passos, L., Vecci, M. A. M., & Young, R. J. (2014). Zoo visitor effect on mammal behavior: Does noise matter? *Applied Animal Behaviour Science*, *156*, 78–84. https://doi.org/10.1016/j.applanim.2014.04.002

Rauscher, F., Robinson, D., & Jens, J. (1998). Improved maze learning through early music exposure in rats. *Neurological Research*, *20*(5), 427–432. https://doi.org/10.1080/01616412.1998.11740543

Reed, H. J., Wilkins, L. J., Austin, S. D., & Gregory, N. G. (1993). The effect of environmental enrichment during rearing on fear reactions and depopulation trauma in adult caged hens. *Applied Animal Behaviour Science*, *36*(1), 39–46. https://doi.org/10.1016/0168-1591(93)90097-9

Robbins, L., & Margulis, S. W. (2014). The effects of auditory enrichment on gorillas. *Zoo Biology*, *33*(3), 197–203. https://doi.org/10.1002/zoo.21127

Robbins, L., & Margulis, S. (2016). Music for the birds: Effects of auditory enrichment on captive bird species. *Zoo Biology* 35(1), 29-34. https://doi.org/10.1002/zoo.21260

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108

Schooler, J. W. (2014). Metascience could rescue the "replication crisis". *Nature*, 515(7525), 9. https://doi.org/10.1038/515009a

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69*, 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An Introduction to registered replication reports at perspectives on psychological science. *Perspectives on Psychological Science*, *9*(5), 552–555. https://doi.org/10.1177/1745691614543974

Singer, H. S. (2011). Stereotypic movement disorders. In W.J. Weiner and E. Tolosa (Eds.). *Handbook of clinical neurology, Vol. 100* (pp. 631–639). Elsevier. https://doi.org/10.1016/B978-0-444-52014-2.00045-8

Smith, M. E., Kane, A. S., & Popper, A. N. (2004). Noise-induced stress response and hearing loss in goldfish (*Carassius auratus*). *Journal of Experimental Biology*, *207*(3), 427–435. https://doi.org/10.1242/jeb.00755

Stoloff, M., Mccarthy, M., Keller, L., Varfolomeeva, V., Lynch, J., Makara, K., Simmons, S., & Smiley, W. (2010). The Undergraduate psychology major: An examination of structure and sequence. *Teaching of Psychology*, *37*, 4–15. https://doi.org/10.1080/00986280903426274

Voellmy, I. K., Purser, J., Flynn, D., Kennedy, P., Simpson, S. D., & Radford, A. N. (2014). Acoustic noise reduces foraging success in two sympatric fish species via different mechanisms. *Animal Behaviour*, *89*, 191–198. https://doi.org/https://doi.org/10.1016/j.anbehav.2013.12.029

Watanabe, S., & Nemoto, M. (1998). Reinforcing property of music in Java sparrows (*Padda oryzivora*). *Behavioural Processes*, *43*(2), 211–218. https://doi.org/10.1016/S0376-6357(98)00014-X

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature*, *582*(7812), 337–340. https://doi.org/10.1038/d41586-020-01751-5

Wells, D. L., & Irwin, R. M. (2008). Auditory stimulation as enrichment for zoo-housed Asian elephants (*Elephas maximus*). *Animal Welfare*, *17*(4), 335–340.

Wells, D. L., Coleman, D., & Challis, M. G. (2006). A note on the effect of auditory stimulation on the behaviour and welfare of zoo-housed gorillas. *Applied Animal Behaviour Science*, *100*(3), 327–332. https://doi.org/https://doi.org/10.1016/j.applanim.2005.12.003

Whipple, B., & Glynn, N. J. (1992). Quantification of the effects of listening to music as a noninvasive method of pain control. *Research and Theory for Nursing Practice*, *6*(1), 43–58.

Wiggins, B. J., & Chrisopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. https://doi.org/10.1037/teo0000137

Williams, I., Hoppitt, W., & Grant, R. (2017). The effect of auditory enrichment, rearing method and social environment on the behavior of zoo-housed psittacines (Aves: *Psittaciformes*); Implications for welfare. *Applied Animal Behaviour Science*, *186*, 85–92. https://doi.org/10.1016/j.applanim.2016.10.013