4-2023

Centring Individual Animals to Improve Research and Citation Practices

Shelly Volsche
Boise State University

Holly Root-Gutteridge
University of Lincoln

Anna T. Korzeniowska
University of Sussex

Alexandra Horowitz
Barnard College

BIOLOGICAL REVIEWS
Cambridge Philosophical Society
Biol. Rev. (2023), **98**, pp. 421–433.
doi: 10.1111/brv.12912

Centring individual animals to improve research and citation practices

Shelly Volsche[1], Holly Root-Gutteridge[2,*], Anna T. Korzeniowska[3,4] and Alexandra Horowitz[5]

[1]*Department of Anthropology, Boise State University, Hemingway Building, 1910 University Drive, Boise, ID, 83725, USA*
[2]*University of Lincoln, Joseph Banks Laboratories, School of Life Sciences, Brayford Pool Campus, Lincoln, LN6 7TS, UK*
[3]*School of Psychology, University of Sussex, Falmer, Brighton, BN1 9QH, UK*
[4]*School of Psychology, University of Surrey, Guildford, GU2 7XH, UK*
[5]*Psychology Department, Barnard College, 3009 Broadway, New York, NY, 10027, USA*

ABSTRACT

Modern behavioural scientists have come to acknowledge that individual animals may respond differently to the same stimuli and that the quality of welfare and lived experience can affect behavioural responses. However, much of the foundational research in behavioural science lacked awareness of the effect of both welfare and individuality on data, bringing their results into question. This oversight is rarely addressed when citing seminal works as their findings are considered crucial to our understanding of animal behaviour. Furthermore, more recent research may reflect this lack of awareness by replication of earlier methods – exacerbating the problem. The purpose of this review is threefold. First, we critique seminal papers in animal behaviour as a model for re-examining past experiments, attending to gaps in knowledge or concern about how welfare may have affected results. Second, we propose a means to cite past and future research in a way that is transparent and conscious of the abovementioned problems. Third, we propose a method of transparent reporting for future behaviour research that (*i*) improves replicability, (*ii*) accounts for individuality of non-human participants, and (*iii*) considers the impact of the animals' welfare on the validity of the science. With this combined approach, we aim both to advance the conversation surrounding behaviour scholarship while also serving to drive open engagement in future science.

Key words: animal behaviour, citation practices, replicability, ecological validity, individual animal.

CONTENTS

* Author for correspondence (Tel.: +441522 83 5297; E-mail: hollyrg@googlemail.com).

I. INTRODUCTION

The scientific investigation of animal behaviour and cognition has undoubtedly brought a wealth of knowledge about how animals perceive the world around them, communicate with each other, obtain food, and learn about their environment. It has enriched our understanding of other species, and about ourselves, through using animals as models for human behaviour, but as we progress in our understanding of animal behaviour and cognition, we learn new and enhanced ways of conducting and reporting studies. Inevitably, questions frequently arise about methods, what scientists do, to whom, and why. While the field of animal behaviour is more than 2000 years old, new findings lead us to constantly revise our approach to data collection and consider the individual animal at the heart of any study.

Contemporary researchers have begun to appreciate that some studies, including seminal and highly cited ones, involved treatment of animals that led to poor outcomes, including physical and psychological trauma. In disregarding welfare, animals were treated as if they were Descartes' 'mere automatons' without feeling or reason (Descartes, 1637, trans. 2004). Early papers often neglected what are now understood as critical roles of animal welfare and ecological validity in animal behaviour experiments, either through not considering them or by not reporting which considerations were made. Many of these studies would be deemed unethical in today's scientific culture, and their methodologies would likely be denied review-board approval. Moreover, by incompleteness or lack of important considerations, they may be, essentially, ecologically invalid. Nonetheless, some of the findings of these research programs became foundational to our understanding of animal behaviour; as such, their results neither can nor should be discarded. For example, Seligman & Meier's (1967) study on learned helplessness used electrical shock and restraint to test dogs' concession to an inescapable, aversive situation. We might imagine researchers today would be hard-pressed to obtain ethical approval for this work, yet according to *Google Scholar*, it has been cited over 3,200 times. In this review, we aim to address the issue of citing fraught research while at the same time acknowledging the welfare, validity, and replicability issues raised, as well as proposing a system to improve methods and reporting in future research. Our aim is to open a conversation about how to reassess past findings under the light of current knowledge and how to cite such research going forward.

At the heart of our proposals is the importance of consideration of the individual animal or animal participants, often missing in reporting of scientific findings (e.g. Thorndike, 2000). This oversight could stem from researchers' struggle with the dilemma presented in the choice between using homogenous populations to reduce bias and error and the potential value of exploring the variation within a species. In many respects, an appreciation for individual animal participants (and the confounds they present) is as important as the control of our variables. As such, even within a controlled study of one species or breed, we argue that individual animals should be recognised as participants of the study.

The significance of the individual participants' role in science reflects that the inquiry of science is not only the search for the causes of each thing (Aristotle, 1984/40 BC), but also a system to train people in the methodology of that search. In the field of animal behaviour, it requires considering the above-mentioned potential sources of variation in the data, and the potential to aggravate or improve the 'replication crisis' (Halina, 2021). The 19th- and early 20th-century papers of many fields fail the basic test of replicability (e.g. Pavlov, 1960), often because they do not provide sufficient information to assess clearly what was done and how. The issues with deficient descriptions in the existing literature can be addressed by building careful assessments of foundational and important papers; partly through considering such work, the means to avoid such pitfalls in future work is revealed. While we focus here specifically on behaviour research using animal models, we suspect many of our ideas could generalise to other research avenues as well.

II. EXAMPLES OF PAST, PROBLEMATIC SCHOLARSHIP

(1) Ivan Pavlov and dogs

Ivan Pavlov was a Russian physiologist studying digestive processes in dogs. His investigations involved measuring salivary production, leading him to identify what we now recognise as classical conditioning. Pavlov and his colleagues spent many years in a purpose-built laboratory in St. Petersburg using hundreds of dogs (exact number unknown; Pavlov, 1960) in various procedures to investigate how classical conditioning works, and, later, which areas of the brain are responsible for certain impulses. Today, virtually all psychology textbooks, if they do mention that Pavlov used dogs as experimental subjects (and not all do; Adams, 2020) refer to Pavlov as the father of classical conditioning, and briefly explain, using simple diagrams, how dogs were conditioned to salivate on hearing a tone (e.g. Klein, 2009).

The reality for Pavlov's dogs, however, was far from this benign and sanitised illustration (Pavlov, 1960). Dogs were subjected to lengthy experiments while strapped to stands. Their salivary glands were surgically exposed and fixed on the outside of their cheeks to measure saliva production

accurately. The stimuli used for the conditioning experiments, apart from typically mentioned tones and flashes, involved strong electric shocks, cuts to the skin, or squirts of acid into the mouth. The dogs underwent multiple surgeries to lesion parts of their brain, which even Pavlov admitted with regret were crude methods (Pavlov, 1960). The procedures left the dogs unable to feed themselves, often blind or deaf, unable to coordinate their movements, or experiencing hyperaesthesia at the slightest touch. Scar tissue formed on the dogs' brains because of imperfect surgeries, leading to dogs suffering from convulsions, and ultimately dying. In some cases, dogs would experience convulsions for many hours (e.g. Pavlov mentions one dog who had fits for 12 h) before dying, without the help of humane euthanasia or anaesthetic (Pavlov, 1960). It is notable, in our considerations of the ethical standards of the early 20th century, that, rather atypically, Pavlov identified his subject dogs by name, including Umnitza, Mampus, and Chingis Khan (Pavlov, 1960).

While Pavlov's methods might have produced groundbreaking findings, they were met with criticism from early on. For example, a Russian baroness Meidenhof, a head of the Central Board of the Russian Society for Animal Protection in 1903 wrote a letter to the War Minister, in which she opposed vivisection [here understood as any experiment on a living organism. It is not entirely clear whether Baroness Meidenhof's criticism was directed at Pavlov specifically or all scientists using animals in experimentation (Kopaladze, 2000)]. However, Pavlov was asked for his opinion and defended vivisection as a regrettable but necessary method employed in pursuit of scientific truth. He also argued that non-specialists should not be interfering in the works of scientists (Kopaladze, 2000).

Further criticism of Pavlov's methods came from outside Russia. For example, in 1909 an English anti-vivisectionist Emilie A.L. Lind-af-Hageby referred to Pavlov's methods as revolting and accused him of lack of understanding for animals, despite acknowledging that Pavlov made an effort to reduce suffering of his experimental subjects as much as possible (Lind-af-Hageby, 1909, cited in Dewsbury, 1990). Similarly, George Bernard Shaw accused Pavlov of ill-treating his experimental subjects, and called his methods 'criminal and detestable' (Shaw, 1947, p. 212, cited in Dewsbury, 1990). More recently, Adams (2020) called for a revision of how psychology describes and reports Pavlov's work. He suggests that literature should include the dogs in the narrative – their stories, their contributions to science and the messy human–animal relationships that had formed between the dogs and the experimenters who worked with them. Perhaps it is time for the sanitised diagrams of salivating dogs to be replaced with a full account of what life was like for Pavlov's dogs?

(2) Harry Harlow and rhesus macaques

Harry Harlow and his colleagues studied various psychological processes in primates, but their research is primarily credited with demonstrating the importance of maternal caregiving to social and cognitive development. Harlow's most famous experiment involved infant rhesus macaques, taken from their mothers within hours of birth, and raised in complete isolation on wire flooring, with only a mother-like figure – either made of wire or covered in terry cloth – for comfort (Harlow & Zimmermann, 1959). Harlow discovered that the infants preferred the contact-comfort of the cloth-covered surrogate 'mother' over food they received from the wire surrogate 'mother' (Harlow & Zimmermann, 1959). Further experiments revealed that the infant macaques, when exposed to fear-inducing stimuli (e.g. a bear-shaped wind-up toy) preferred cloth-covered mothers for comfort over the wire mothers (Harlow & Zimmermann, 1959).

Harlow & Zimmerman (1958) concluded that bodily contact with a mother-like figure is essential for the development of affection and love for their mothers in infants, trumping the importance of nutrition received from a mother figure. This finding sparked numerous studies in attachment in humans and other species (Klein, 2009) leading to Harlow's papers being cited hundreds of times, often without any acknowledgement of the considerable cost for the rhesus macaque subjects. These included trauma of early separation from their mothers, isolated rearing (highly unnatural for a social species) on barren flooring, and being subjected to fear-inducing stimuli, as well as open field tests used to assess the monkeys' behaviour in the absence of their surrogate mothers. As expected by Harlow and colleagues, the infants used their cloth mothers as 'safe havens' when scared, but in some experiments they were deprived of these comfort figures – leading to clear signs of extreme emotional distress such as 'crouching, rocking and sucking', 'frantic clutching of their bodies', clutching of a cotton diaper (Harlow & Zimmermann, 1959, p. 505) and 'screaming in abject terror' (Harlow & Zimmermann, 1959, p. 423). Harlow's legacy also includes subjecting infant monkeys to months of isolation incarcerated within metal chambers, dubbed 'pits of despair' – an experiment aimed at creating a state akin to human depression (characterised by low locomotion and exploration, high incidence of self-clasping and huddling behaviour, and lack of interest in social interactions) from which the monkeys never recovered (Suomi & Harlow, 1972). Harlow's work, and in particular the treatment of his experimental subjects, attracted much criticism over the years, both from the public as well as researchers working in Harlow's laboratory (Remele, 2018). Some questioned the purpose of the experiments given there were already existing studies on maternal separation in human infants (e.g. by John Bowlby and Rene Spitz; Remele, 2018). Many did not believe the costs of his studies in terms of animal life were offset by the gains in scientific discovery (Stephens, 1986) and some simply called the studies out as sadistic (Haraway, 1989). Nonetheless, Harlow's papers are still highly cited and praised without mention of the subjects' poor welfare, the effects therefrom or ethical issues surrounding the experiments (see Gluck, 1997). Furthermore, the work started by Harlow continues and creates new controversies. For example, in 2015 PETA alerted

the US House of Representatives to the ethically dubious experiments being conducted by Stephen Suomi, Harlow's former doctoral student and collaborator, on infant rhesus monkeys. While the experimental procedures have been adjusted to address some of the concerns, not everyone is satisfied that the work is even necessary (Reardon, 2015).

(3) Martin Seligman and dogs (and other species)

Martin Seligman and colleagues were interested in how depression in humans arises and studied it experimentally using dogs (and later other species). Seligman & Meier (1967) decided to study the effect of inescapable painful events on dogs' subsequent behaviour. They placed 30 mixed-breed dogs in sound-proof cubicles and strapped them into rubberised hammocks which restrained their heads and left their feet dangling. Brass-plated electrodes were attached to the dogs' hind footpads, through which 64 electric shocks of six milliamps (considered painful to humans) were delivered, for up to 120 s each. One group of dogs could switch off the shock by pressing a panel on the side of their heads. The dogs in the 'inescapable' condition had a non-functional panel that did not discontinue the shock. As a result of this treatment, when the dogs were placed into a shuttle box delivering electric shocks of up to 60 s and a barrier to jump for safety, the 'inescapable' dogs did not make any efforts to avoid being shocked. Rather, they passively endured the pain for the duration. By contrast, the dogs who could eliminate the shock by pressing a panel during the first experiment quickly jumped over the shuttle box barrier and into the safe zone. This inability to control the environment was thought to develop helplessness in the dogs assigned to the 'inescapable' condition, a state which Seligman directly compared with depression in humans (Seligman & Meier, 1967). Over the years scientists replicated Seligman and Meier's results with other non-human species [e.g. Masserman (1971) in cats or Frumkin & Kenneth (1969) in fish] and humans (Hiroto, 1974; using noise instead of electric shocks as a noxious stimulus). However, over the years the validity of learnt helplessness as a model for depression started to be questioned as too simplistic and underplaying the heterogeneous nature of depression in humans (e.g. Buchwald, Coyne & Cole, 1978) eventually prompting Seligman and colleagues to reformulate their model of depression, recognising that the laboratory animal model could not be used to explain the aetiology of clinical depression in humans after all (Abramson, Seligman & Teasdale, 1978).

(4) Why are the studies problematic?

The subjects used in these examples suffered poor welfare, including unsuitable living conditions, confinement, painful procedures, and psychological trauma. This is problematic considering what we now know about non-human animal sentience (e.g. Sneddon, 2019; Valenchon *et al.*, 2017), the ability to feel pain (e.g. Jeong *et al.*, 2020) and experience psychological distress (e.g. Poole, 1997; Asiedu *et al.*, 2021). Methods that result in poor animal welfare are not only distressing, but they also threaten to invalidate the scientific claims of these studies. It is, after all, now widely recognised that poor physical and psychological welfare affects the results of studies using animal models (see Henke, 1997). For example, the unresponsiveness of Pavlov's dogs may be less related to architecture of their cerebral cortex and instead related to experiencing post-operative pain (Pavlov, 1960).

A further problem relates to the ecological validity of the studies and their generalisability to humans. For example, while Harlow's studies were popular and led to future studies of attachment in humans, it can be argued that methodology of these studies (severe psychological trauma; experimental set ups incongruent with human infants' experiences) invalidates the results and generalisability to humans. An additional concern is the underreporting of the animal outcomes in the experiments: how many were used (e.g. Pavlov's studies), housing conditions [e.g. some information on housing is available for Pavlov (1960), and Seligman & Meier (1967), but very little for Harlow & Zimmermann (1958)], and endpoint care (e.g. euthanasia, adoption; information missing in most of the aforementioned studies). These examples serve to inform our exposition, below, of types of critical methodological details that exemplar studies could – but may or may not have – reported.

III. REVISING CITATION PROTOCOL

Given the advancements in scientific understanding over the decades, and the discovery of problematic methods in research, we propose that questioning research in general citations should be *de rigueur*. If research is simply cited as 'Pavlov (1960)' in passing, the effect and import of poor husbandry and invasive medical procedures is lost. While it is often impossible to explore the nuanced findings and issues of every citation, the collective effort of post-publication 'review' can help contextualise and identify transgressions. A mechanism for citing research studies that are now seen as problematic or flawed is essential.

We propose highlighting citations of research with the consistent issues of missing data or poor descriptions of data collection, ecological or methodological validity concerns, and welfare concerns. Importantly, we do not suggest that these research programs and resultant papers not be cited at all. Particularly for transformative scholarship, we acknowledge that neglecting findings is not advisable when building new models or developing future frameworks. In fact, we know that underreporting of null findings is as prevalent an issue as the replication crisis (see Franco, Malhotra & Simonovits, 2014). Additionally, ignoring problematic literature negates the opportunity for professional learning and growth in the field. Instead, we suggest that innovative but ethically or methodologically problematic scholarship be

cited but also identified as such. In this way, the culture of science remains responsive to past flaws and future refinement and continues the dialogue such that we do not repeat past mistakes.

There are increasing calls for accountability and contextualisation of research, both of which require transparency in the presentation of data (e.g. Cojocaru & Von Gall, 2019). Currently, the convention is to cite the earliest work that produced a conclusion or demonstrated a fact, as well as recent literature which has refined that work. This leads to a chain of citations that do not critically examine the original research nor the potential issues. For example, mentions of Pavlov rarely add that the dogs' salivary glands were externalised (e.g. Harlow & Zimmermann, 1958, as cited by Alpher, 1984) almost certainly a welfare concern that possibly affected his results. Thus, problems in earlier research can be lost in the noise of their significance to the field. More modern research is expected to adhere to broader ethical frameworks established and revised over decades [including the recommendations of the National Centre for the Replacement, Refinement, and Reduction of Animals in Research (NC3R) to replace animals when possible, refine studies, and reduce the number of animals needed] as well as local ethical review board criteria (although even within political bodies there can be variations in what is required for ethical approval). However, much past research would not meet the current guidelines for ethics either due to poor reporting (deficient description) or welfare concerns, making it difficult to assess whether the results obtained were reliable. The result is false confidence that foundational, highly cited research must be robust, despite evidence that this may not be true.

(1) A proposed system for marking research as 'problematic' in citations

Any system that attempts to address the issue of marking all potentially problematic research with a single reference is bound to fail, or even to encompass the complexity of issues that can occur. However, here we attempt to begin the conversation and suggest a route forward. We considered several proposals for achieving this goal (see Table 1). First, one might use footnotes – a strategy common in books, but often not permitted in journals. Another possibility is a description in the body of the text demonstrating awareness of the problems of work cited. However, this may be unwieldy and burdensome, given the strict word limits placed by many journals. A third possibility is giving cited research a numerical 'score': a rating of the validity, completeness of data, and adherence to accepted approaches such as the guidelines for the complete and ethical reporting of *in vivo* animal research laid out in the NC3R or Animal Research: Reporting of *In Vivo* Experiments (ARRIVE) guidelines. This approach, however, would not enable the elaboration of the problem or problems seen – and is also more definitive than, we suggest, authors can be about the degree of the problems. Importantly, too, scientific understanding underpinning ideas of ecological validity and the standards for welfare and ethics will no doubt continue to evolve, and any highlighting of these will be entirely contemporary judgements, themselves likely to become outdated within a few years of publication. A fourth option imagines that asterisks or other markers – previously common, for instance, in dictionaries that marked words which were considered substandard (Merriam-Webster, 1993; Heritage, 1969) – could be used. Without context or knowledge of their meaning, though, these markers could leave readers uninformed. Of course, authors could also decide not to include the citations at all, treating knowledge gained by unethical means as *pro tanto* morally wrong (Tuvel, 2015), but this approach is also fraught. Human medical researchers have debated the ethics of citing experiments that were not simply unethical but are bywords for horror, in particular the work of Nazi scientists (Moe, 1984; Cohen, 1990). A broader conversation on how to treat knowledge gained under unethical circumstances is required before such a far-reaching approach can be adopted.

Here, we do not conclude that science is best served by rejecting the citing of past experiments that would fail contemporary ethical standards. We propose, instead, that in the first instance a single marker word, '*problematic*', be applied following a citation thus: '(Pavlov, 1960: *problematic*)'. Use of such a marker directs the reader to the reference to consider whether they wish to reassess the information cited, refer directly to the paper and assess for themselves the issues the author detected, or accept that the data presented in it may be compromised or incomplete, but remain important enough to cite. This system has the benefit of being brief and to the point. It does, of course, lack nuance and the very contextualisation for which we argue. Therefore, we additionally suggest the following criteria markers for assessing papers:

(1) *Problematic: Deficient Description* (DD) – This addition would highlight papers that lack detailed descriptions of the methodology, animal husbandry, validity, or other considerations that allow readers to vet the work accurately. For example, metrics for monitoring stress, housing conditions, pain relief options, social isolation, opportunities for animals to dissent, and criteria for inclusion may not be listed. As a result, the research may not be replicable from the details given.
(2) *Problematic: Welfare Validity Concerns* (WVC) – This addition highlights research results in which animals' responses were likely affected by pain, hunger and thirst, fear and distress, inability to express species-appropriate behaviour, or housing conditions. It may be that increasing knowledge about a species changes our perception of good welfare for them, thus this is not a label to accuse researchers of maltreatment of their subjects. Instead, new information can alter our interpretation of past results. This issue arises for invertebrates as well as vertebrates, for example it is now accepted that fish feel pain (Sneddon, 2019) which can affect their behaviour (Deakin *et al.*, 2019) and snails are affected by stress (Lukowiak *et al.*, 2014): good science will rely on good animal welfare (Poole, 1997).

Table 1. Potential solutions to marking problematic research in citations.

Option	Description	Pros	Cons
Asterisks	* Marks questionable content	Short Will not affect word count Can be integrated into citation	Non-descriptive May not fit with journal formatting
Scores	Numerical ratings	Short Will not affect word count Can be integrated into citation	Non-descriptive May not fit with journal formatting Requires judgement for scores Subjectivity/Potential Bias
Footnotes	Issues noted in footnote	Strong descriptive power Allows discussion of issues May not count for word count (journal dependent)	Long May count against word count (journal dependent) May not fit with journal formatting
Paragraphs	Full paragraph in text describing the issues	Strong descriptive power Allows discussion of issues in depth and with nuance Fits into journal formatting	Long Adds to word count Unwieldy for multiple citations Balancing relevance to overall paper
'Problematic'	Single word used to mark issues in the work	Short Has little impact on word count Can be integrated into citation Can be expanded with Deficient Description (DD), Ecological Validity Concerns (EVC), or Welfare Validity Concerns (WVC) markers	Limited descriptive power May not fit with current journal formatting
Do not cite	Authors could choose not to include problematic papers	No impact on word count, citations Avoids the ethical dilemma	May incite reviewer requests for inclusion May be difficult to establish theoretical framework of some research Places some knowledge out of bounds Subject to contemporary standards for 'knowledge gained from evil' Potential for censoring the past

(3) *Problematic: Ecological Validity Concerns* (EVC) – This addition would be useful if the current or expanded understanding of the species' ecology is not acknowledged in the cited research. Lacking ecological validity does not mean that the animal necessarily suffered or was distressed, though it certainly overlaps with WVC quite frequently. Specifically, EVC indicates that some aspect of the studied animal's known behaviour was overlooked, whether because of researcher unfamiliarity or a failure of imagination. An example of this is the finding that the pain inhibition of laboratory mice is affected by male but not female experimenters and thus baseline readings of their physiology are affected (Sorge *et al.*, 2014): any previous research results including measures of pain in mice are now rendered questionable on the basis of this finding.

Naturally, these proposals raise the question of who should levy judgement on past research. While it is tempting to suggest a globalised database of agreed-upon ratings, that is far beyond the scope of our current capacity to create – although perhaps a goal for the future. We suggest that individual authors should assess papers for themselves, train future scholars in this practice, and use the terms when they have the knowledge required properly to assess the work. Reviewers may also recommend the use of these terms when appropriate. As is common in the review process, we anticipate a discussion would unfold between reviewers, authors, and editors, further achieving our goal of encouraging discussion and critique as part of peer review. Regardless of the approach, this system should be applied to all research, not just classic or foundational works, and concerns should be noted whenever they occur.

Ideally, individual journals and writing style guides would begin to normalise one of these approaches and include it in the author guidelines for submission preparation. This would create a sort of shared responsibility for adherence and reporting. There is already precedent for including ethical documentation in many journals, and prominent journals

like *Animal Behaviour* have specific requirements to include an ethical note and document how research methods and housing align with the PREPARE (Planning Research and Experimental Procedures on Animals) checklist for excellence in animal research and the ARRIVE guidelines. Both PREPARE and ARRIVE require a higher ethical and reporting standard than most institutional review boards (Smith *et al.*, 2018; Percie du Sert *et al.*, 2020).

IV. PROPOSAL FOR A STANDARDISED, TRANSPARENT METHOD FOR REPORTING FUTURE RESEARCH

The second part of our two-part proposal is forward-looking. We suggest a multipronged approach to ensure methodological transparency in future research. Revising expectations for research is not a new concept: ethics surrounding research in any field evolve to incorporate updated knowledge, new sociocultural norms, and technological advances in methods – including the importance of considering good animal welfare as integral to good laboratory science (Gluck, 2016). Good welfare depends on knowledge of both the individual and the species in question, as we detail below. While it is likely impossible to 'future proof' science against validity concerns, by including a thorough and transparent presentation of details regarding the animals and their treatment, it is possible to avoid at least some of the pitfalls of deficient descriptions.

(1) Changing perspectives on ethical approval

Ethical review for human subjects research is constantly revised to account for new research or legal considerations. For example, The Belmont Report of 1979 (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979) outlined specific expectations related to human subjects research, stipulating 'that individuals should be treated as autonomous agents and … that persons with diminished autonomy are entitled to protection' (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, part B1). Modern researchers such as Sebo (2022) argue that it is time to take these steps with non-humans, too. Furthermore, Ferdowsian *et al.* (2020) specifically outline how to apply the ethical principles of The Belmont Report to non-human animals, drawing comparisons between children and non-human animals as equally vulnerable populations in need of protection. Ferdowsian *et al.* (2022) identify the structural changes necessary to encourage the cultural shift necessary for more ethical and just approaches, and further suggest how to incentivise these choices through tenure policies, funding reviews, and institutional processes. Our proposal adds consideration of the *umwelt* of individual animals as an integral part of ethical review and research design. Without this knowledge, we may not appropriately recognise the protections needed beyond the freedom from hunger and thirst; from discomfort; from pain, injury, or disease; from fear and distress; and to express normal behaviour (collectively known as the five freedoms). An integrated approach considers the study aims, the individual animal's experience of the experiment, the severity of the proposed experiment, and the species' general welfare requirements. We suggest that the result will be better science.

Creativity in methods and exhaustive literature reviews can help researchers avoid problematic work (DD, WVC, or EVC). Failure to go above the minimum legal expectations in our work with animals results in a waste of lives, time, and money, and can perpetuate the perception that replicating restrictive animal models is our only option. This is the case despite Wurbel's (2000) highlighting of the standardisation fallacy in behaviour research's generalisability outside of controlled environments and Voelkl *et al.*'s (2020) efforts to incorporate systematic heterogenisation of sample populations to reflect better the influence of environment conditions on genotypes and behavioural expression. This may be even more critical for invertebrate taxa, the ethical standards for whom are evolving quickly, and which often lack legal protection, making individual researchers' choices more important (Drinkwater, Robinson & Hart, 2019).

While most researchers already undergo various forms of internal ethical and legal review, there is a lack of consistency among countries, institutions, and ethical review board members. Moreover, the advancement of the protection of animals has been uneven across taxa. For example, the US Government passed the Animal Welfare Act of 1966 (AWA) to define and regulate treatment of livestock, research animals, and companion animals (pets). In 1976, the definition of 'animal' was expanded to specify primarily warm-blooded, vertebrate species as protected (Animal Welfare Act Amendments of 1976: Congress, 1976). Yet within that definition, exceptions are made. The working definition of 'animal' in US legal arenas is still centred largely around companion animals, identifies certain species as exempt based upon breeding and use (e.g. homogenous laboratory strains), and provides little to no protection for livestock, even in research environments. Most ethical review boards in the US are based upon the guidelines set forth in the AWA, and often refrain from making ethical judgments beyond legal mandates. They are also seated with individuals who themselves benefit from animal research (Hansen, 2013). By contrast, the UK's Animal (Scientific Procedures) Act 1986 and the European Parliament's Directive 2010/63/EU protect any vertebrate or cephalopod species, including independently feeding larvae and mammalian foetuses in the third trimester of development. Further highlighting the inconsistency, all species of non-human animals are protected under Japanese legislation (Kurosawa, 2007), while the South African Animals Protection Act of 1962 emphasises domestic animals (including birds) and wildlife in captivity or under control of any person with no mention of research or scientific oversight.

While some legislation provides a foundation for ethics review, one could argue (e.g. Woodruff, 2019) that updates

are rarely sufficient in responding to changes in the field of behaviour and cognition research – especially for ectotherms and non-charismatic species such as rodents. Moreover, some legislation actually harms animals, as an 'obstacle to transparency' (Marceau, 2018, p. 925). Although improvements are taking place in many parts of Europe (e.g. UK, Sweden, Austria; World Animal Protection, Retrieved May 19, 2022), the majority of nations tracked by World Animal Protection are either similar or less protective in terms of animal treatment and protection. Therefore, we argue that it is essential that researchers begin to hold ourselves and each other accountable for the quality of this preliminary work. Here again, the journal *Animal Behaviour* might serve as an example towards such efforts, as it publishes animal welfare and ethics expectations for all submissions to be considered (Guidelines for the treatment of animals in behavioural research and teaching, 2020).

(2) Honouring individual animals in research

Implicit in our proposals is the acknowledgement of the importance of the individual subject animals' lives. In the experiments mentioned above, the animals were virtually invisible as individuals – apart from providing data such as response rates or drops of saliva. Who they were is rarely mentioned despite the fact that without their contribution the resulting scientific discoveries would not have been possible. How can we put the animal in the foreground of animal behaviour research, recognising their critical role in science and their individuality? One possibility is by naming the animals (Table 2). This is rarely done in the literature and is obviously not always possible – as when dealing with schools of fish or very large populations of rodents. Yet it seems something must be done to move us from treating them as interchangeable objects in scientific inquiry to individual subjects in our laboratories and the world. One exception is more recent research into dog cognition, where the individuals are named by their owners, but it need not be restricted to privately owned animals. By listing names, one allows future researchers to track individuals' progress across multiple studies. For instance, Alex the African grey parrot, studied for decades by Irene Pepperberg (Pepperberg, 2002, 2006, 2007), was named in research; thus, it was possible to track his individual progress as he learned new words. Additionally, it was understood that his personality and life history may have influenced his learning abilities. With an increasing exploration of how cognition may vary tremendously between species members (Fugazza *et al.*, 2021; Kaminski, Call & Fischer, 2004), individual identification has become more important to interpret research results and names can permit the tracking of an individual's participation across publications.

(3) Assent, dissent, and the role of individual choice

In animal behaviour, the importance of individual choice has been increasingly recognised. Researchers thus may consider adopting an assent–dissent model of participation (Kantin & Wendler, 2015). Dissent is used here to describe an expressed behavioural objection to a condition (e.g. freezing, crying out, or otherwise expressing discomfort; Kantin & Wendler, 2015). The concepts of dissent and assent are often assessed at a species level. However, individuals of a species may show the same capacity for the experiment but show individual preference for participation, influencing results.

As an example of this: two dogs are given the same task of learning to wear headphones and hold a pose for a functional magnetic resonance imaging (fMRI) experiment. Zen happily does so, offering the behaviour, but Lucy resists. By using only positive reinforcement and assent model training, only dogs who are willing to participate in the fMRI scans are included (e.g. Berns, Brocks & Spivak, 2013). Similarly, Pepperberg's testing proceeded only when Alex was willing: thus, assent was not only considered at an individual level but was regularly requested and assessed (Pepperberg & Carey, 2012; Pepperberg & Gordon, 2005).

A useful feature of using the assent–dissent model (Kantin & Wendler, 2015) is that species-specific moral judgements do not apply. A flaw is that the assessment of assent depends on the researcher's ability to translate the behaviour of the animal correctly, which often relies on extensive time and experience with the model species. This is a challenge that will require the engagement of all behavioural researchers and ask for their individual-level choices to be reflected in their reports. Additionally, researchers should always report individuals who dissent in order to capture and present the variation within species or groups.

We recognise that there are more welfare concerns outside the moment of research study. For example, housing is often inadequate, and acquisition of research subjects may not always be ethical. The ways in which we unpack ecological validity, welfare validity, and data deficiency earlier in this review continue to be relevant in the development and reporting of new research, too.

(4) Future-proofing and avoiding deficient descriptions

Our proposal builds from Hooijmans, Leenaars & Ritskes-Hoitinga (2010) calling for a 'gold standard in reporting' in systematic reviews to assess better the living conditions of subjects in laboratory animal experiments, and also draws from Kilkenny *et al.*'s (2010) ARRIVE Guidelines and Webster & Rutz's (2020) STRANGE framework. (The STRANGE framework ensures that subjects are representative by considering their Social background; Trappability and self-selection; Rearing history; Acclimation and habituation; Natural changes in responsiveness; Genetic make-up; and Experience.) These papers highlighted the need to report housing, enrichment, nutrition, population variation, handling, and the importance of framing good welfare as good research, not just an ethical and legal requirement, and laid out systems for doing so. While their checklist was focused specifically on laboratory animal protocols, and was more descriptive than prescriptive, their review

Table 2. Suggested format for reporting data in future and assessing data from the past.

Variable type	Variable	Example 1 – Deficient description (Pavlov, 1960)	Example 2 – Ecological/methodological validity issues (Suomi & Harlow, 1972)	Example 3 – Replicable (Pepperberg, 1983)
Demographic data	Species – common name	Dog	Rhesus monkeys	African grey parrot
	Species – Latin name	Not stated	*Macaca mulatta*	Not stated
	Sample size	Unknown	Four experimental + four controls	One
	Sex of individuals	Unknown	Male	Male
	Age (mean + SD)	Unknown	45 days at the start of experiment	13 months old
	Names or identifying labels (where applicable)	Some names mentioned in the reporting e.g. 'Brains' (but mostly anonymous)	No names mentioned	Alex
	Body mass (where applicable)	No information	No information	No information
	Inclusion/ exclusion criteria	No information	No information	No information
Husbandry	Source of animals	No information	None provided (likely born in the laboratory)	Purchased in a pet shop in the Chicago area
	Housing for animals	Kennels, dogs kept in groups	Experimental subjects: first 45 days in laboratory nursery, then in bare ¾ inch wire-mesh cages 2.5 × 2 × 2.5 foot from 90 days of age; Controls: first 30 days in laboratory nursery, then in cages as above either on their own (2 controls) or with another equal-aged male partner (2 controls)	Allowed free access to all parts of laboratory 8 h/day on request and during sleeping hours housed in a parrot cage 62 × 62 × 73 cm with water and food available *ad libitum*
	Enrichment provided	No information	None provided	Fresh fruit, vegetables, nuts and toys available on vocal request
	Endpoint for animal/ mortality rate	Some dogs died as a result of procedures	No information	No information
Experimental details	Experimental outcome measures	Salivary secretions measured in drops of saliva	Behavioural observations in experimental chambers, home cages and playroom sessions (with other monkeys present) – mean scores for various behaviour categories recorded (e.g. self-groom, play, huddle, etc.)	Accuracy of vocal labelling of items presented to the parrot as a % of questions asked during the testing session

(Continues on next page)

Table 2. *(Cont.)*

Variable type	Variable	Example 1 – Deficient description (Pavlov, 1960)	Example 2 – Ecological/methodological validity issues (Suomi & Harlow, 1972)	Example 3 – Replicable (Pepperberg, 1983)
	Procedure/ experiment type	Experimentation involving classical conditioning to various stimuli – lesioning of parts of the brain ('partial destruction or complete extirpation') followed by experimentation as above; dogs' salivary glands surgically exposed and tubes or bulbs fixed to them to collect saliva; dogs set up in stands and restrained with loops during experimentation	Experimental subjects confined in a vertical, stainless-steel chamber at 45 days of age until 90 days of age. Subjects able to eat and drink and move freely in three dimensions. At 4 months of age subjected to 6 30-min playroom sessions/week where they had access to the control subjects	Following training to acquire names of categories relating to the colour, shape, or materials of an object the parrot was tested by presentation of exemplars of items followed by questions related to specific attributes of the items e.g. 'What colour?'; correct responses were rewarded by being allowed to interact with desirable items and incorrect responses were negatively punished by withholding the toy
	Justification for protocol	The use of secretory reflexes allowed extremely accurate measurements of the intensity of the reflex activity; it did not lend itself to anthropomorphic interpretation; experimental environment was finely controlled by separating the experimenter from the dog to eliminate any confounding stimuli that could disrupt the conditioning process; the method of brain lesioning justified as 'the only method so far available' for the 'study of cortical localisation of functions' (Pavlov, 1960, p. 320)	Separation from others used previously to model human anaclitic depression resulting in identical behaviours in human and monkey infants. Vertical chamber designed on an intuitive basis to reproduce physically and psychologically a depression-like state ('a well of despair'). Suomi and Harlow's previous study indicated that a chamber produces depression-like behaviours in monkeys. Chamber used here to explore further its effectiveness in producing psychopathology.	A human model/rival creates competition for the reward of interacting with an object, as well as the trainer's attention; this avoids the need for external rewards while enabling the possibility of revealing latent abstract conceptual understanding.
	Welfare monitoring scheme (with/ without citation)	No formal monitoring but dogs that were ill/had no appetite were temporarily withdrawn from experiments and treated	No information	No information
Participation criteria	Termination criteria for trial	Dogs refusing food or being dull and unresponsive; dog falling asleep in stand	No information	No information
	Termination criteria for participation in study/exit point	No information	No information	No information

(Continues on next page)

Table 2. *(Cont.)*

Variable type	Variable	Example 1 – Deficient description (Pavlov, 1960)	Example 2 – Ecological/methodological validity issues (Suomi & Harlow, 1972)	Example 3 – Replicable (Pepperberg, 1983)
	Criteria for animal's assent or dissent to participate	No information	No information	No formal acknowledgement but information on signs that Alex lost interest in testing (indicated by turning his back to the trainer, requesting corks or clothes pins)
Administrative	Ethics approval reference	None provided	None provided	None provided
	Data access	None provided	None provided	None provided
	Funders	None provided	None provided	None provided
	Notes	None provided	None provided	None provided

provided extensive detail on the impact that welfare can have on science (Hooijmans *et al.*, 2010). We suggest that an elaborated table (Table 2) be included in all future animal behaviour papers as supplementary material. The table groups several important variables into general categories: 'Demographic data' and 'Husbandry' are rather self-explanatory and fit the details required by ARRIVE 2.0 (Percie du Sert *et al.*, 2020). 'Experimental details' is not intended to reiterate the methods section but to prompt lucid, brief descriptions of the approach used for quick reference. 'Participation criteria' allows for numeration of which demographic details were considered in selection and how the animals' assent or dissent was measured (see Section IV.3). Finally, 'Administrative' covers various record-keeping details.

To demonstrate the potential utility of this approach, we completed Table 2 for three classic papers which exemplify issues with (*i*) deficient descriptions; (*ii*) ecological validity; and (*iii*) a demonstration of a more thorough approach to ecological validity considerations, welfare, and data reporting. Much of the information requested in Table 2 is routinely presented in good research reports now. Still, standardising the reporting structure and proposing its use before ethical review would greatly improve the quality of work – as well consider individual subjects.

As proposed, this table can be used for the ongoing peer review of existing literature, by prompting questions about how it was performed. It may encourage researchers to reassess their knowledge by standardising the questions asked and could potentially be used to teach students how to perform robust peer reviews of research. By doing so during the research design phase, the table also becomes a tool for training, a means of checking ethical standards, and a preparatory document that can assist in filing ethical review applications and protocols.

V. DISCUSSION

As researchers, we should remain aware at all times of the context in which research was done and its aims. Indeed, far from dismissing or 'cancelling' the research that falls short of these standards, we argue that only by recognising where it faltered can we avoid similar pitfalls. Replicability, the underpinning of any scientific endeavour, requires both transparency and detailed descriptions. Perfection is not an achievable condition in science: we will always operate from a flawed position of incomplete knowledge and control. In this way, science is a self-improving, iterative endeavour as the knowledge of past missteps informs how to improve. This iterative process makes challenging accepted evidence more important, especially as we come to understand the relevance of different variables on behaviour.

We argue that considering animal welfare in research does not merely raise ethical or philosophical questions. While harm and stress to the participant animals should be avoided

whenever possible, their presence should be considered variables within the experiment and listed as such. Without the transparency of listing housing conditions, welfare metrics, and enrichment efforts, it would not be possible to assess whether animals' performances were influenced by these factors. We should also not assume there are universals across species: what negatively affects one species may not be a negative for another (e.g. Carlstead & Brown, 2005; Schapiro *et al.*, 1996).

VI. CONCLUSIONS

(1) In this review, we have critiqued problematic reporting in foundational research and proposed two important changes in the citation and reporting of past and future work.
(2) First, we proposed a means to cite fundamentally valuable papers in a way that is transparent and conscious of data, ethical, and validity concerns.
(3) Second, we proposed a means of transparent reporting for future behaviour research that (*i*) improves replicability, (*ii*) accounts for individual non-human participants, and (*iii*) considers the impact of the animals' welfare on the validity of the science. We recognise that examples used herein represent a bias toward larger animal species. This is a product of our own research emphases and suggests a broader bias in animal behaviour research.
(4) Ultimately, our goal is to begin a conversation that centres animals as participants and subjects in our work and acknowledges the value of proper welfare, not solely for the animals' wellbeing, but also for the continual refinement of science. We call upon researchers, journal editors, and ethical review boards to embark upon the next phase of animal welfare, a transition that will improve our work and the wellbeing of the species we study.

VII. ACKNOWLEDGEMENTS

The authors would like to thank Jeff Sebo and Sara Waller for their helpful commentary on earlier drafts and the anonymous reviewers whose suggestions improved the paper. We would also like to thank our dogs, Calvin, Emma, Finnegan, Lucy, Max, Quiddity, Sheba, and Upton, for their assent to participation in our research and for the insight into other species' *umwelt* that we have gained from them, helping to inform this paper. This work was supported by the Culture & Animals Foundation.

VIII. REFERENCES

Abramson, L. Y., Seligman, M. E. & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology* **87**(1), 49–74.
Adams, M. (2020). The kingdom of dogs: understanding Pavlov's experiments as human–animal relationships. *Theory and Psychology* **30**(1), 121–141.
Alpher, V. S. (1984). Attachment: a reconsideration of its heterogeneity and construct validity. *Psychological Reports* **54**(2), 531–535.
Aristotle (1984). *The Complete Works of Aristotle: The Revised Oxford Translation*. Bollingen Series. Princeton University Press, Princeton.
Asiedu, E., Karlan, D., Lambon-Quayefio, M. & Udry, C. (2021). A call for structured ethics appendices in social science papers. *Proceedings of the National Academy of Sciences* **118**, e2024570118.
Berns, G. S., Brooks, A. & Spivak, M. (2013). Replicability and heterogeneity of awake unrestrained canine fMRI responses. *PLoS One* **8**(12), e81698.
Buchwald, A. M., Coyne, J. C. & Cole, C. S. (1978). A critical evaluation of the learned helplessness model of depression. *Journal of Abnormal Psychology* **87**, 180–193.
Carlstead, K. & Brown, J. L. (2005). Relationships between patterns of fecal corticoid excretion and behavior, reproduction, and environmental factors in captive black (*Diceros bicornis*) and white (*Ceratotherium simum*) rhinoceros. *Zoo Biology* **24**, 215–232.
Cohen, B. (1990). The ethics of using medical data from Nazi experiments. *Journal of Halacha and Contemporary Society* **19**, 103–126.
Cojocaru, M.-D. & Von Gall, P. (2019). Beyond plausibility checks: a case for moral doubt in review processes of animal experimentation. In *Animal Experimentation: Working Towards a Paradigm Change*, pp. 289–304. Brill, Leiden.
Congress, 94th (1976). *Animal Welfare Act Amendments of 1976*. United States Government Print Office. United States.
Deakin, A. G., Buckley, J., AlZu'bi, H. S., Cossins, A. R., Spencer, J. W., Al'Nuaimy, W., Young, I. S., Thomson, J. S. & Sneddon, L. U. (2019). Automated monitoring of behaviour in zebrafish after invasive procedures. *Scientific Reports* **9**, 9042.
Descartes, R. 1637, trans. (2004). *A Discourse on Method: Meditations and Principles*. Translated by Veitch, John. Orion Publishing Group, London.
Dewsbury, D. A. (1990). Early interactions between animal psychologists and animal activists and the founding of the APA committee on precautions in animal experimentation. *American Psychologist* **45**, 315–327.
Drinkwater, E., Robinson, E. J. H. & Hart, A. G. (2019). Keeping invertebrate research ethical in a landscape of shifting public opinion. *Methods in Ecology and Evolution* **10**, 1265–1273.
Ferdowsian, H., Fuentes, A., Johnson, L. S. M., King, B. J. & Pierce, J. (2022). Toward an anti-Maleficent research agenda. *Cambridge Quarterly of Healthcare Ethics* **31**, 54–58.
Ferdowsian, H., Johnson, L. S. M., Johnson, J., Fenton, A., Shriver, A., Gluck, J., Syd, L., Johnson, M., Johnson, J., Fenton, A., Shriver, A. & Gluck, J. (2020). A Belmont report for animals? *Cambridge Quarterly of Healthcare Ethics* **29**, 19–37.
Franco, A., Malhotra, N. & Simonovits, G. (2014). Publication bias in the social sciences: unlocking the file drawer. *Science* **345**, 1502–1505.
Frumkin, K. & Kenneth, H. (1969). Conditioned fear training and later avoidance learning in the goldfish. *Psychonomic Science* **16**, 159–160.
Fugazza, C., Andics, A., Magyari, L., Dror, S., Zempléni, A. & Miklósi, Á. (2021). Rapid learning of object names in dogs. *Scientific Reports* **11**, 2222.
Gluck, J. P. (1997). Harry F. Harlow and animal research: reflection on the ethical paradox. *Ethics and Behavior* **7**, 149–161.
Gluck, J. P. (2016). *Voracious Science & Vulnerable Animals: A Primate Scientist's Ethical Journey*. University of Chicago Press, Chicago, IL.
Guidelines for the treatment of animals in behavioural research and teaching (2020). *Animal Behaviour* **159**, I–XI.
Halina, M. (2021). Replications in comparative psychology. *Animal Behavior and Cognition* **8**, 263–272.
Hansen, L. A. (2013). Institution animal care and use committees need greater ethical diversity. *Journal of Medical Ethics* **39**, 188–190.
Haraway, D. J. (1989). *Primate Visions: Gender, Race, and Nature in the World of Modern Science*. Routledge, New York and London.
Harlow, H. F. & Zimmermann, R. R. (1958). The development of affectional responses in infant monkeys. *Proceedings of the American Philosophical Society* **102**, 501–509.
Harlow, H. F. & Zimmermann, R. R. (1959). Affectional response in the infant monkey. *Science* **130**, 421–432.
Henke, S. (1997). Do white-tailed deer react to the dinner bell? An experiment in classical conditioning. *Wildlife Society Bulletin* **25**, 291–295.
Heritage, A. (1969). *The American Heritage Dictionary of the English Language*. Houghton Mifflin Harcourt Trade & Reference Publishers, New York.
Hiroto, D. S. (1974). Locus of control and learned helplessness. *Journal of Experimental Psychology* **102**, 187–193.
Hooijmans, C. R., Leenaars, M. & Ritskes-Hoitinga, M. (2010). A gold standard publication checklist to improve the quality of animal studies, to fully integrate the three Rs, and to make systematic reviews more feasible. *Alternatives to Laboratory Animals* **38**, 167–182.
Jeong, M.-J., Lee, C., Sung, K., Jung, J. H., Pyo, J. H. & Kim, J.-H. (2020). Fear response-based prediction for stress susceptibility to PTSD-like phenotypes. *Molecular Brain* **2020**(13), 1–9.

Kaminski, J., Call, J. & Fischer, J. (2004). Word learning in a domestic dog: evidence for 'fast mapping'. *Science* **304**, 1682–1683.
Kantin, H. & Wendler, D. (2015). Is there a role for assent or dissent in animal research? *Cambridge Quarterly of Healthcare Ethics* **24**, 459–472.
Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biology* **8**, e1000412.
Klein, S. B. (2009). *Learning: Principles and Applications*, Fifth Edition. Sage Publications, Inc, Thousand Oaks.
Kopaladze, R. A. (2000). Ivan P. Pavlov's view on vivisection. *Integrative Physiological and Behavioural Science* **35**, 266–271.
Kurosawa, T. M. (2007). *Japanese Regulation of Laboratory Animal Care with 3Rs*, Edition (Volume **14**), pp. 317–321. Japanese Society for Alternatives to Animal Experiments, Tokyo.
Lukowiak, K., Sunada, H., Teskey, M., Lukowiak, K., Dalesman, S., Davies, S. A., Dow, J. A. T. & Lukowiak, K. (2014). Environmentally relevant stressors alter memory formation in the pond snail *Lymnaea*. *Journal of Experimental Biology* **217**, 76–83.
Marceau, J. (2018). How the animal welfare act harms animals. *Hastings Law Journal* **69**, 925–959.
Masserman, J. H. (1971). The principle of undcertainty in neurotigenesis. In *Experimental Psychopathology: Recent Research and Theory* (ed. H. D. Kimmel). New York Academic Press, New York.
Merriam-Webster (1993). *Merriam-Webster's Collegiate Dictionary*, Tenth Edition. Globe Fearon Educational Publishing, New York.
Moe, K. (1984). Should the Nazi research data be cited? *Hastings Center Report* **14**, 5–7.
National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979). *The Belmont Report 18 April 1979*. US Department of Health, and Human Services, Washington. Electronic file available at https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html.
Pavlov, I. P. (1960). *Conditioned Reflex: An Investigation of the Physiological Activity of the Cerebral Cortex*. Dover Publications, Oxford.
Pepperberg, I. M. (1983). Cognition in the African Grey parrot: preliminary evidence for auditory/vocal comprehension of the class concept. *Animal Learning & Behavior* **11**, 179–185.
Pepperberg, I. M. (2002). *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*. Harvard University Press, Cambridge.
Pepperberg, I. M. (2006). Cognitive and communicative abilities of Grey parrots. *Applied Animal Behaviour Science* **100**(1–2), 77–86.
Pepperberg, I. M. (2007). Grey parrots do not always 'parrot': the roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Sciences* **29**, 1–13.
Pepperberg, I. M. & Carey, S. (2012). Grey parrot number acquisition: the inference of cardinal value from ordinal position on the numeral list. *Cognition* **125**, 219–232.
Pepperberg, I. M. & Gordon, J. D. (2005). Number comprehension by a Grey parrot (*Psittacus erithacus*), including a zero-like concept. *Journal of Comparative Psychology* **119**, 197–209.
Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., Browne, W. J., Clark, A., Cuthill, I. C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S. T., Howells, D. W., Karp, N. A., *et al.* (2020). The ARRIVE guidelines 2.0: updated guidelines for reporting animal research*. *Journal of Cerebral Blood Flow and Metabolism* **40**, 1769–1777.
Poole, T. (1997). Happy animals make good science. *Laboratory Animals* **31**(2), 116–124.
Reardon, S. (2015). NIH defends monkey experiments. *Nature*. 1–2. doi:10:1038/nature2015.16814
Remele, K. (2018). When Harry meets Harry: an ethical assessment of Harry Harlow's maternal deprivation experiments. In *The Ethical Case against Animal Experiments* (eds A. Linzey and C. Linzey), pp. 200–205. University of Illinois Press, Champaign.
Schapiro, S. J., Bloomsmith, M. A., Porter, L. M. & Suarez, S. A. (1996). Enrichment effects on rhesus monkeys successively housed singly, in pairs, and in groups. *Applied Animal Behaviour Science* **48**, 159–171.
Sebo, J. (2022). *Saving Animals, Saving Ourselves: Why Animals Matter for Pandemics, Climate Change, and Other Catastrophes*. Oxford University Press, Oxford.
Seligman, M. E. P. & Meier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology* **74**, 1–9.
Smith, A., Clutton, R. E., Lilley, E., Hansen, K. E. A. & Brattelid, T. (2018). PREPARE: guidelines for planning animal research and testing. *Laboratory Animals* **52**(2), 135–141.
Sneddon, L. U. (2019). Evolution of nociception and pain: evidence from fish models. *Philosophical Transactions of the Royal Society B* **374**(1785), 20190290.
Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., Rosen, S., Tuttle, A. H., Wieskopf, J. S., Acland, E. L., Dokova, A., Kadoura, B., Leger, P., Mapplebeck, J. C. S., McPhail, M., Delaney, A., Wigerblad, G., *et al.* (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods* **11**, 629–632.
Stephens, M. L. (1986). *Maternal Deprivation Experiments in Psychology: A Critique of Animal Models*. American Antivivisection Society, Jenkingtown.
Suomi, S. J. & Harlow, H. F. (1972). Depressive behavior in young monkeys subjected to vertical chamber confinement. *Journal of Comparative and Physiological Psychology* **80**, 11–18.
Thorndike, E. L. (2000). *Animal Intelligence: Experimental Studies*, First Edition. Routledge, New York.
Tuvel, R. (2015). Against the use of knowledge gained from animal experimentation. *Societies* **5**(1), 220–244.
Valenchon, M., Lévy, F., Moussu, C. & Lansade, L. (2017). Stress affects instrumental learning based on positive or negative reinforcement in interaction with personality in domestic horses. *PLoS One* **12**, e0170783.
Voelkl, B., Altman, N. S., Forsman, A., Forstmeier, W., Gurevitch, J., Jaric, I., Karp, N. A., Kas, M. J., Schielzeth, H., van de Casteele, T. & Würbel, H. (2020). Reproducibility of animal research in light of biological variation. *Nature Reviews Neuroscience* **21**, 384–393.
Webster, M. M. & Rutz, C. (2020). How STRANGE are your study animals? *Nature* **582**(7812), 337–340.
Woodruff, M. L. (2019). Smart sheep need more protection. *Animal Sentience* **4**, 9.
World Animal Protection, 2022. *Animal protection index*. Retrieved May 19. Electronic file available at https://api.worldanimalprotection.org/
Wurbel, H. (2000). Behaviour and the standardization fallacy. *Nature Genetics* **26**, 263.

(*Received 25 May 2022; revised 5 October 2022; accepted 6 October 2022; published online 25 October 2022*)