Technical contribution: a cautionary note on the use of behavioural diversity (H-Index) in animal welfare science

KA Cronin*† and SR Ross‡

† Animal Welfare Science Program, Lincoln Park Zoo, 2001 North Clark Street, Chicago, IL 60614, USA
‡ Lester E Fisher Center for the Study & Conservation of Apes, Lincoln Park Zoo, 2001 North Clark Street, Chicago, IL 60614, USA
* Contact for correspondence and requests for reprints: kcronin@lpzoo.org

Abstract

Animal welfare scientists actively seek reliable and practical metrics that can serve as indicators of animal welfare for use with agricultural, laboratory and zoo-housed animals. Behavioural diversity as a welfare concept originated from early welfare scientists linking poor animal welfare with a high proportion of time spent engaged in stereotyped behaviours and little expression of behaviours that would be expected under wild, or natural conditions. Recently, the concept of behavioural diversity as a welfare indicator has been widely adopted and is frequently quantified by the Shannon-Wiener Diversity Index (H-index). However, the H-Index is fraught with theoretical and mathematical limitations when applied to animal welfare. Four key problems with this metric are demonstrated: the metric's responsivity to the size of the behavioural repertoire and the underlying assumption that larger behavioural repertoires reflect better welfare, the sensitivity of the metric to arbitrary decisions about the resolution of behavioural categories, the fact that the calculation of the metric is agnostic to the valence of behaviours, and the metric's susceptibility to the common practice of excluding some behavioural categories. Moving forward, we recommend focusing on validated welfare measures that are sensitive to valence when evaluating animal welfare.

Keywords: *animal welfare, behavioural diversity, H-index, Shannon-Wiener Diversity Index, welfare metrics, well-being*

Introduction

Scientists actively seek reliable and practical metrics that can serve as indicators of animal welfare. While several definitions of animal welfare exist, the core component to many definitions is the subjective experience of the animal or their affective state (Keeling *et al* 2011). Given that the affective experience of an animal cannot be easily measured, welfare assessments often rely on behavioural measures that may provide insight into the subjective experience of the animal. While the number of validated, behavioural indicators of welfare is still limited (Mason & Mendl 1993; Dawkins 2015), there are some specific behaviours which are increasingly regarded as reflective of an animal's affective state, both positive and negative (ie Mason & Latham 2004; Mellor 2015). Recently, there has been increasing interest in the use of a compound measure of welfare, behavioural diversity, which quantifies the scope of behaviours being expressed, and proportion of time invested in different behaviours. Here, we question the reliability of this compound behavioural measure as a welfare indicator.

Interest in behavioural diversity as a measure of welfare can be traced to the work of Alex Stolba and others to explain the behaviour, and particularly the stereotyped behaviour, of domestic pigs (*Sus scrofa*) living in impoverished environments. Stolba *et al* (1983) coined the phrase "informational redundancy" to explain the emergence of these stereotypies and hypothesised that a function of stereotypies might be to achieve a homeostasis in arousal by avoiding external stimuli and creating sensory inputs that are more predicable in nature. As a greater proportion of the animal's behavioural repertoire was expressed as these redundant stereotyped behaviours, the overall breadth of natural behaviour correspondingly decreased. This thinking was further supported by studies showing that while sows living in semi-natural conditions exhibited more than 100 different behavioural patterns, this number decreased severely to just 33 in sows living in stalls (Stolba *et al* 1983). A similar narrowing of behavioural diversity to achieve homeostasis in arousal was seen even more dramatically when stereotypes were pharmacologically induced in laboratory rats (*Rattus norvegicus*) injected with

amphetamines. Following the injection, complex behaviours disappeared first and were replaced by fewer behaviours with shorter durations, and ultimately only simple tremor behaviour remained (Lyon & Robbins 1975).

The inverse relationship between the proportion of time an animal engages in stereotyped behaviour and the overall diversity of behaviours appears to have driven the development of behavioural diversity as a potential welfare metric. Furthering this was an associated growth in interest in promoting natural patterns of behaviour (Hancocks 1980; Fabregas *et al* 2012, but see Veasey *et al* 1996; Fraser 2008). A functional perspective of behaviour proposes that animals have evolved a range of behavioural strategies which help them maximise survival and reproduction and that environments which prevent the expression of that full range of behaviour result in motivational discord and behavioural frustration (Duncan & Petherick 1991; Fraser & Duncan 1998). One outcome of this 'natural living' perspective is the potential adoption of behavioural diversity as a means of measuring the degree to which animals are given all the behavioural opportunities that would be available to them in the complex environment in which they evolved. Simply put, in this perspective, animals that are only expressing a fraction of the behaviours demonstrated by their wild counterparts are behaviourally restricted and may be experiencing frustration and diminished welfare. Following this line of thought, some welfare scientists appear to have made the assumption that more behaviours expressed equates to more naturalistic states, and therefore better welfare (Collins *et al* 2016; Miller *et al* 2016; but see Vickery & Mason 2004). It is this line of reasoning that draws our critical interest.

The predominant strategy for quantifying behavioural diversity is calculation of the Shannon Diversity Index (H), also referred to as the Shannon-Wiener H-index (Shannon & Weaver 1949), a method originally developed to quantify uncertainty in strings of text (Shannon 1948). When used to quantify behavioural diversity, this measure is impacted by the number of different behaviours (or behavioural categories) that are expressed, and how evenly represented the different behaviours are in the repertoire. The H-index is calculated as follows:

$$H = - \sum_{i=1}^{R} (p_i \ln p_i)$$

Where R = the full set of behavioural categories expressed by the animal during a period of observation, and p_i = the proportion of time engaged in the ith behavioural category. A higher value of H indicates greater behavioural diversity and, as such, better welfare.

Interpreting the H-index as an indicator of welfare has recently gained popularity in a variety of applied ethological studies in zoos, laboratories and to a lesser extent, agri-cultural settings, and with a broad range of taxa (eg armadillos [*Dasypus novemcinctus*], sloths [*Galago senegalensis*] and galagos [*Choloepus didactylus*]: Clark & Melfi 2012; cheetahs [*Acinonyx jubatus*]: Miller *et al* 2016; chimpanzees [*Pan troglodytes*]: Neal Webb *et al* 2018; cichlids [*Oreochromis mossambicus*]: Galhardo *et al* 2008; dolphins [*Tursiops truncatus*]: Miller *et al* 2011; felids [*Felis* spp]: Shepherdson *et al* 1993; foxes [*Vulpes vulpes*]: Kistler *et al* 2009; giraffes [*Giraffa camelopardalis*]: Razal *et al* 2017; gorillas [*Gorilla gorilla*]: Charmoy *et al* 2015; penguins [*Pygoscelis papua*]: Collins *et al* 2016; pigs: Hirt & Wechsler 1994 and Wemelsfelder *et al* 2000; squirrel monkeys [*Saimiri sciureus*]: Izzo *et al* 2011; wolves [*Canis lupus*]: Frezard & Pape 2003; multiple species, reviewed in Brown *et al* 2006). Given the surge in publications relying heavily on the H-index to interpret welfare status of a wide range of species, a careful evaluation of the utility of this metric seems warranted. We provide here an evaluation of how applicable the use of the H-index is to the assessment of captive animal welfare and where the promise of such an approach may fall short. By identifying the strengths and weaknesses of this behavioural research method, we hope efforts to improve the lives of captive animals with empirical approaches can be enhanced.

Materials and methods

We have identified four primary limitations associated with the use of the H-index to evaluate behavioural diversity and, by extension, animal welfare. The shortcomings relate to both the input, the assumptions researchers make about behaviours prior to calculating the index, and to the output, the way the index is then interpreted. In the sections that follow, we have opted to organise our criticism of the application of the H-index in welfare science around four main flaws.

Flaw 1 — the assumption that bigger is better

By definition, in cases where behavioural categories are equally expressed, the H-index increases asymptotically as more behavioural categories are added to the repertoire (Figure 1). As a greater H-index is interpreted as better welfare, this leads to scenarios whereby bigger behavioural repertoires are interpreted as reflecting better welfare. This assumed positive relationship between the number of behavioural categories and welfare may have originally emerged from observations that animals often exhibited a smaller behavioural repertoire in captive environments compared with the same species under natural conditions (see above). While the expression of natural behaviour likely has strong ties to welfare (Farm Animal Welfare Council 1992; Bracke & Hopster 2006; Brando & Buchanan-Smith 2018), a blind assumption that a larger behavioural repertoire reflects better welfare is problematic. While, in many cases, it is possible that more behaviours reflect better welfare, there are several realistic scenarios under which this relationship does not hold, such as when behaviours indicative of poor welfare are added to the repertoire.

Figure 1
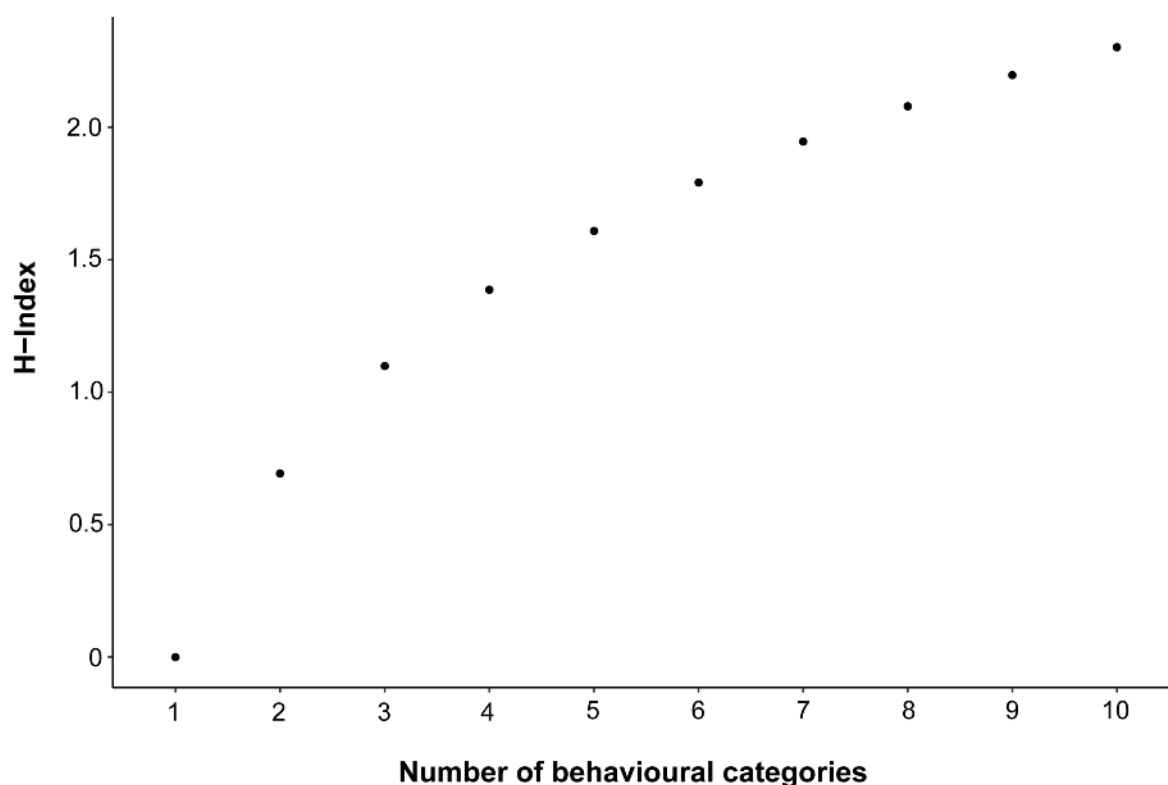
The relationship between the H-Index and the number of behavioural categories. Under the condition that the behavioural categories are equally expressed (p1 = p2 = …pi), the H-index is equal to the natural log of the number of behavioural categories.

To provide an example, consider a scenario in which a laboratory colony of mice (*Mus musculus*) previously living in opaque, polypropylene caging have been introduced to new caging with stainless steel wire bars that allow visual access outside the cage. The colony previously did not display any vigilance behaviours, but upon moving to the new environment, the mice spend increasingly large proportions of their time in an anxious state monitoring the environment outside the cage. In the case that this new behaviour occurs at a high enough frequency to equate to proportions of time spent in the other, pre-existing behaviours, a higher H-index could be obtained in the new environment simply because a novel behaviour has been introduced. The H-index would indicate improved welfare as the number of behavioural categories has increased, whereas closer inspection of the components of the repertoire may be far less optimistic about the impact of this broadened behavioural repertoire.

Flaw 2 — the H-index is impacted by behavioural splitting and lumping

A second, related limitation of the H-index emerges from the fact that different H-indices could be obtained depending upon the way behavioural categories are generated from the specific behaviours present in an ethogram. The behavioural categories that are used to calculate the H-index tend to be quite coarse and include broad categories such as, locomotion, feeding, social behaviour, abnormal behaviour, and inactivity. However, most researchers utilise ethograms in practice that have much more granularity (eg social behaviours on the ethogram may include aggressive behaviour, grooming or preening, sexual behaviours, etc). The decision on how to split or lump behaviours in the ethogram into behavioural categories for analyses is not straightforward, and two observers watching the same animal at the same time could result in different H-indices simply because different behavioural categories are selected (eg one observer decides to use a catch-all category of 'social behaviour' and another creates two categories, 'affiliative social behaviours' and 'aggressive social behaviours'). These hypothetical scenarios are intended to demonstrate that one who tended to lump behaviours into fewer categories could generate a different H-index than one who tended to split behaviours into more categories (Table 1). Given that there is no standard way to create behavioural categories (even within species), the responsivity of the H-index to this arbitrary decision has the potential to be a problematic issue if the adoption of the H-index continues to grow and comparisons are attempted between studies.

Table 1 Resolution of behavioural categories impacts H-index.

Original set of behavioural categories		Set of behavioural categories with increased resolution	
Behavioural category	Proportion of time	Behavioural category	Proportion of time
A	0.1	a1	0.05
		a2	0.05
B	0.1	b1	0.05
		b2	0.05
C	0.1	c1	0.05
		c2	0.05
D	0.1	d1	0.05
		d2	0.05
E	0.1	e1	0.05
		e2	0.05
F	0.1	f1	0.05
		f2	0.05
G	0.1	g1	0.05
		g2	0.05
H	0.1	h1	0.05
		h2	0.05
I	0.1	i1	0.05
		i2	0.05
J	0.1	j1	0.05
		j2	0.05

A hypothetical set of behavioural categories that are performed for an equal proportion of time is shown on the left. On the right is a second possible set of behavioural categories and corresponding proportion of time when each category has been split into two. The higher resolution categories shown on the right would result in a higher H-index (H = 3.0 compared H = 2.3), although the behavioural repertoire and activity budget could be identical. These side-by-side tables demonstrate how the arbitrary splitting or lumping of behavioural categories generates different H-indices when underlying behavioural profiles have not changed.

Flaw 3 — the assumption that all behaviours were created equal

The predominant measure used to characterise behavioural diversity, the H-index, equally weights all behavioural categories, regardless of what they are. As researchers are equating greater behavioural diversity (higher H-index) with better animal welfare, they are by extension assuming that all behavioural categories in a repertoire have equal ability to contribute to positive welfare. This assumption is not appropriate because we

know that some behaviours have been clearly demonstrated to be associated with better welfare whilst others are associated with poorer welfare. For example, monkeys administered anxiogenic drugs spend a greater proportion of time self-scratching, yawning and engaged in locomotor stereotypies (Major *et al* 2009), providing a clear link between negative welfare (affective state) and these specific behaviours. Examples of such links to demonstrate positive welfare are rarer but behaviours such as social play have been shown to be indicative of positive welfare states (eg Brown *et al* 2015), and our ability to objectively determine the valence of behaviours is continually advancing. For example, recent work has relied on endocrinology to infer positive valence of grooming and sharing food among chimpanzees (Wittig *et al* 2014) and vampire bats (*Desmodus rotundus*) (Carter & Wilkinson 2015), and emerging methods to measure affect through cognitive biases shed light on the valence of specific expressed behaviours (eg dogs [*Canis familiaris*]: Mendl *et al* 2010; rats: Rygula *et al* 2012; capuchin monkeys (*Cebus apella*): Pomerantz *et al* 2012). Behaviours do have valence in the context of welfare, however, this valence is not considered in the H-index since all behaviours are treated similarly.

We include a final hypothetical scenario to illustrate this problem. Consider a situation in which a welfare scientist is attempting to use the H-index to evaluate welfare using five behavioural categories. Figure 2(a) shows the five behaviours occurring at equal proportions, maximising the H-index that could be obtained with this number of behavioural categories (H = 1.61). Figure 2(b) shows the five behaviours occurring at different proportions that arguably reflect a more positive welfare state because the animal spends less time engaged in aggressive interactions and more time engaged in playful behaviour, yet the H-index associated with B is lower (H = 1.56), quantitatively indicating poorer welfare.

Flaw 4 — the H-Index does not know what you have not told it

A number of authors have attempted to deal with the inherent difference in valence across behaviours by *a priori* excluding behaviours from the H-index calculation. One common strategy is to exclude behaviours that are not species-typical, or that are known to be associated with poor welfare (eg Miller *et al* 2016). However, this approach can also generate misleading results. Consider a hypothetical study in which researchers collect behavioural data that can be logically grouped into five behavioural categories: play; sleep; feed; aggression; locomotion; and stereotypic. The researchers are interested in whether a certain treatment impacts animal welfare as measured by the H-index, and they exclude the category 'stereotypic' from their calculations. It is entirely possible that the H-index indicates an increase in behavioural diversity following the treatment while, in reality, there was a substantial increase in stereotypic

Figure 2

Demonstration that the H-index is not responsive to behavioural changes typically associated with improved welfare. The graphs show the proportion of time an animal spent engaged in each of five behavioural categories. In (a) the animal spent an equal proportion of time engaged in all five behavioural categories, meanwhile in (b) the animal spent less time engaged in aggression and more time engaged in play. The H-index would be higher for the behavioural profile shown in (a) (H_a = 0.2[ln(0.2)]+ 0.2[ln(0.2)]+ 0.2[ln(0.2)] + 0.2[ln(0.2)]+ 0.2[ln(0.2)] = 1.61), but the content of the behaviours strongly suggests better welfare for the behavioral profile shown in (b) (H_b = 0.3[ln(0.3)] + 0.2[ln(0.2)] + 0.2[ln(0.2)] + 0.1(ln[0.1]) + 0.2[ln(0.2)] = 1.56).
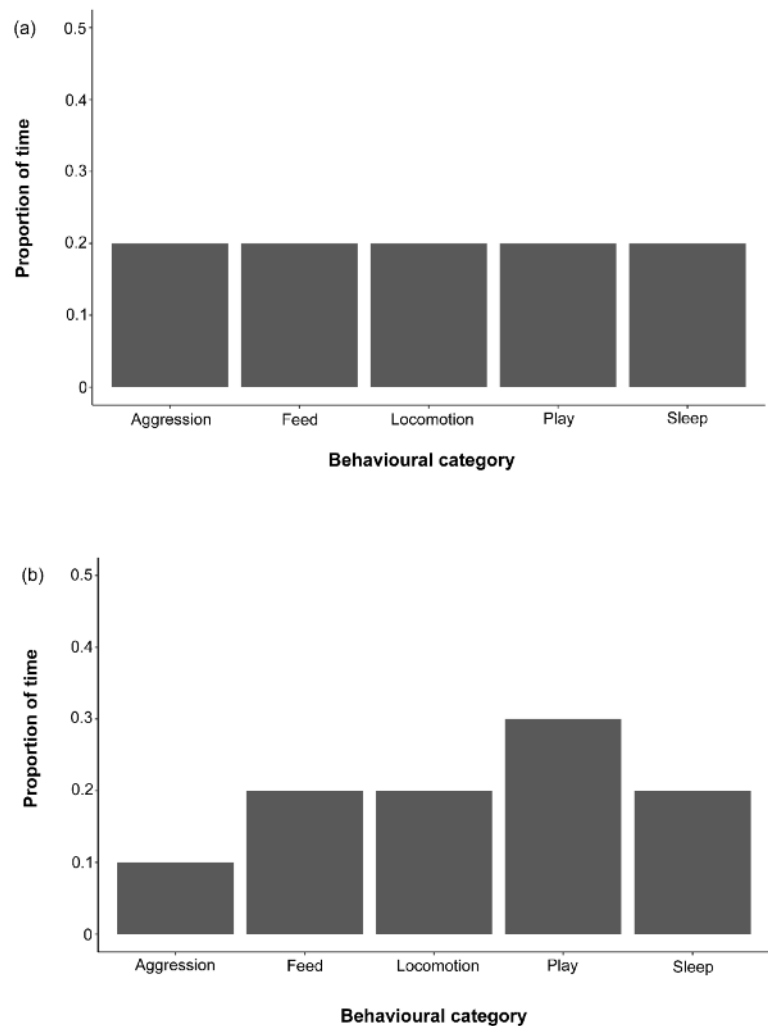
behaviour reflective of poor welfare (Figure 3). The above hypothetical scenario highlights how a study designed to evaluate welfare can generate problematic conclusions about effective ways to improve welfare by relying on the H-index. This example represents a plausible scenario as some researchers are moving away from reporting changes in activity budgets and relying solely on reports of the H-index.

Discussion

The analysis of behaviour as a means to monitor, assess and ultimately improve animal welfare has a long history in agricultural, laboratory and zoological environments (Broom 1988; Gonyou 1994; Dawkins 2004). While the majority of research has focused on behavioural indicators of negative welfare states, such as stereotypies (ie Mason & Latham 2004), there is a growing impetus to identify behaviours or sets of behaviours that would signal positive states of welfare (eg Boissy *et al* 2007; Napolitano *et al* 2009; Mellor 2015,

2016). As part of these efforts, the adoption of a composite measure of behaviour, behavioural diversity, which conveys the degree to which a presumably healthy breadth of behaviours is expressed, has been increasingly utilised. Here, we have outlined some of the limitations of the use of the standard behavioural diversity index (H) to guide future uses, interpretation and refinement of compound behavioural measures.

We remain cautiously optimistic that another measure that encompasses not only the breadth of expressed behaviours but also carefully considers the valence of the behaviour as it relates to welfare could be useful. However, the use of the H-index as a tool to assess welfare has fundamental, mathematical limitations that result in findings that can be at best difficult to interpret and at worst, misleading. Applied ethologists have demonstrated that single, composite welfare measures can be problematic because underlying value assumptions are unavoidable (Tannenbaum 1991; Fraser 1995). If researchers are determined to have a composite

Figure 3
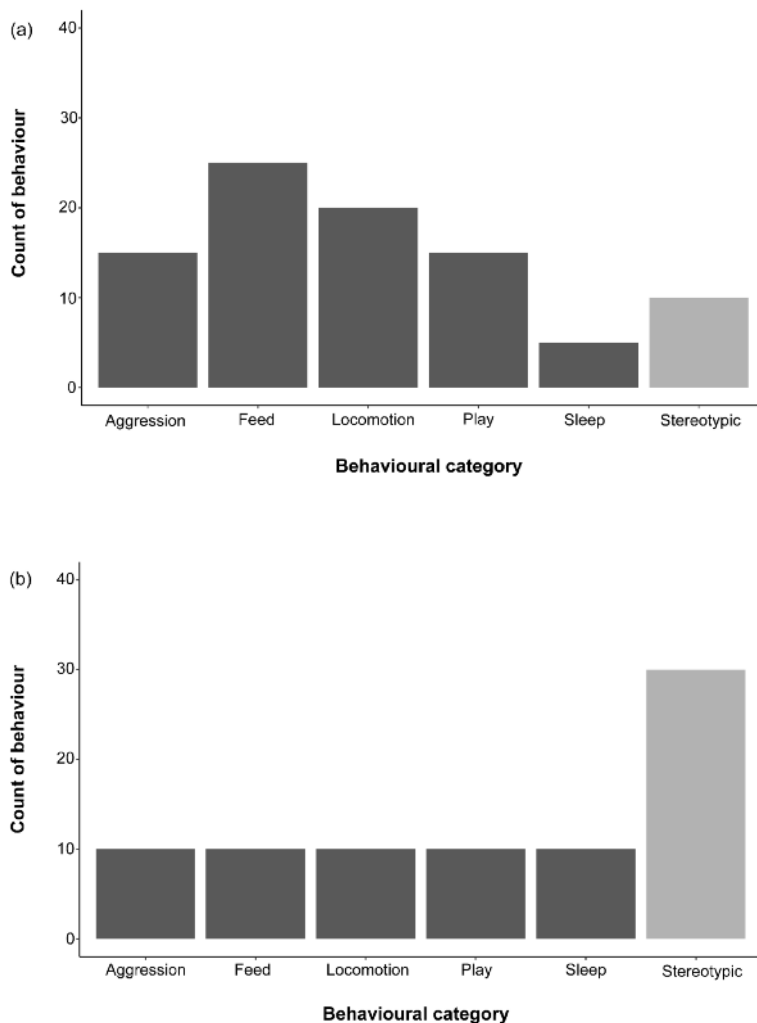
Demonstration that the common practice of excluding behavioural categories can generate unreliable H-index values. Figure shows the counts of behaviours observed during the same amount of observation time for (a) pre- and (b) post-treatment. The count of stereotypical behaviours increases post-treatment (shown in light grey). If the category stereotypic behaviour is excluded from the H-index, the H-index indicates increases, suggesting an improvement in welfare post-treatment.

index to apply to welfare assessment, it should not be one that is agnostic with regards to valence. Furthermore, it should be one that is transparent about the individual behavioural changes that have occurred, rather than simply reporting an overall change in the behavioural diversity index (see Shepherdson *et al* 1993). We encourage an approach that identifies and targets changes in specific behaviours that are likely to reflect welfare rather than a compound behavioural measure.

The term behavioural diversity is used across a variety of contexts, many of which have little to do with individual welfare. The importance of maintaining a broad behavioural repertoire across generations of animals, especially those destined for reintroduction back into their native habitats, is an important aspect of behavioural management and conservation, but one that does not relate directly to our discussions of individual welfare. Rabin (2003) suggests that such conservation-minded approaches to behavioural diversity may even conflict with approaches that target individual welfare. Additionally, those studying social learning also employ the term behavioural diversity when discussing variation in behaviour across individuals or populations (Griffin *et al* 2014). The establishment of behavioural diversity as a key concept in other related fields may have facilitated its hasty adoption in the field of welfare science.

By considering the fundamental limitations of the most common approach to measuring behavioural diversity (H), and the inherent difficulties in relating increases in H-index to legitimate enhancements of animal welfare, we hope to have illustrated that much work remains to be done before investing in this single composite measure. An approach that encompasses specific, measurable and reliable associations with internal affective states would be an improvement, and work is continually progressing in a direction that may make this feasible in the future. Currently, we should resist the widespread adoption of the H-index as it is an imperfect measure that may lead to a false sense of assuredness that we have developed a means to measure, and ultimately improve, the welfare of captive animals.

© *2019 Universities Federation for Animal Welfare*

Animal welfare implications

There is widespread interest in development of positive welfare indicators. One potential positive indicator that has been employed by welfare scientists working in a range of environments is behavioural diversity. However, here, we show that behavioural diversity, especially as quantified by the H-index, is an unreliable measure of welfare. Despite its lack of validity, this measure is currently being used to evaluate the welfare impact of husbandry and management practices for animals in a variety of settings. By calling attention to the potential of this measure to lead to inaccurate conclusions about changes in welfare status, we hope to minimise the chances that practices are adopted that may negatively impact animal welfare.

Acknowledgements

References

Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, Winckler C, Forkman B, Dimitrov I, Langbein J and Bakken M 2007 Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior 92*: 375-397. https://doi.org/10.1016/j.physbeh.2007.02.003

Bracke MB and Hopster H 2006 Assessing the importance of natural behavior for animal welfare. *Journal of Agricultural and Environmental Ethics 19*: 77-89. https://doi.org/10.1007/s10806-005-4493-7

Brando S and Buchanan-Smith HM The 24/7 approach to promoting optimal welfare for captive wild animals. *Behavioural Processes 156*: 83-95. https://doi.org/10.1016/j.beproc.2017.09.010

Broom DM 1988 The scientific assessment of animal welfare. *Applied Animal Behaviour Science 20*: 5-19. https://doi.org/10.1016/0168-1591(88)90122-0

Brown M, Carbone L, Conlee KM, Dawkins MS, Duncan IJ, Fraser D, Griffin G, Hampshire VA, Lambert LA, Mench JA and Morton D 2006 Report of the working group on animal distress in the laboratory. *Lab Animal 35*: 26. https://doi.org/10.1038/laban0906-26

Brown SM, Klaffenbock M, Nevison IM and Lawrence AB 2015 Evidence for litter differences in play behaviour in pre-weaned pigs. *Applied Animal Behaviour Science 172*: 17-25. https://doi.org/10.1016/j.applanim.2015.09.007

Carter GG and Wilkinson GS 2015 Intranasal oxytocin increases social grooming and food sharing in the common vampire bat *Desmodus rotundus*. *Hormones and Behaviour 75*: 150-153. https://doi.org/10.1016/j.yhbeh.2015.10.006

Charmoy K, Sullivan T and Miller LJ 2015 Impact of different forms of environmental enrichment on foraging and activity levels in gorillas, *Gorilla gorilla gorilla*. *Animal Behavior and Cognition 2*: 233-240. https://doi.org/10.12966/abc.08.03.2015

Clark FE and Melfi VA 2012 Environmental enrichment for a mixed-species nocturnal mammal exhibit. *Zoo Biology 31*: 397-413. https://doi.org/10.1002/zoo.20380

Collins C, Quirke T, Overy L, Flannery K and O'Riordan R 2016 The effect of the zoo setting on the behavioural diversity of captive gentoo penguins and the implications for their educational potential. *Journal of Zoo and Aquarium Research 4*: 85-90

Dawkins MS 2004 Using behaviour to assess animal welfare. *Animal Welfare 13*: S3-S7

Dawkins MS 2015 Animal welfare and the paradox of animal consciousness. *Advances in the Study of Behavior 47*: 5-38. https://doi.org/10.1016/bs.asb.2014.11.001

Duncan IJ and Petherick JC 1991 The implications of cognitive processes for animal welfare. *Journal of Animal Science 691*: 5017-5022. https://doi.org/10.2527/1991.69125017x

Fàbregas MC, Guillén-Salazar F and Garcés-Narro C 2012 Do naturalistic enclosures provide suitable environments for zoo animals? *Zoo Biology 31*: 362-373. https://doi.org/10.1002/zoo.20404

Farm Animal Welfare Council 1992 *The Five Freedoms*. Farm Animal Welfare Council: London, UK

Fraser D 1995 Science values and animal welfare: exploring the 'inextricable connection' *Animal Welfare 4*: 103-117

Fraser D 2008 Toward a global perspective on farm animal welfare. *Applied Animal Behaviour Science 113*: 330-339. https://doi.org/10.1016/j.applanim.2008.01.011

Fraser D and Duncan IJ 1998 Pleasures, pains, and animal welfare: toward a natural history of affect. *Animal Welfare 7*: 383-396

Frézard A and Pape GL 2003 Contribution to the welfare of captive wolves *Canis lupus lupus*: a behavioral comparison of six wolf packs. *Zoo Biology 22*: 33-44. https://doi.org/10.1002/zoo.10070

Galhardo L, Correia J and Oliveira RF 2008 The effect of substrate availability on behavioural and physiological indicators of welfare in the African cichlid *Oreochromis mossambicus*. *Animal Welfare 17*: 239-254

Gonyou HW 1994 Why the study of animal behavior is associated with the animal welfare issue. *Journal of Animal Science 72*: 2171-2177. https://doi.org/10.2527/1994.7282171x

Griffin AS, Diquelou M and Perea M 2014 Innovative problem solving in birds: a key role of motor diversity. *Animal Behaviour 92*: 221-227. https://doi.org/10.1016/j.anbehav.2014.04.009

Hancocks D 1980 Bringing nature into the zoo: inexpensive solutions for zoo environments. *International Journal for the Study of Animal Problems 13*: 170-177

Hirt H and Wechsler B 1994 Behavioural diversity as a measure of welfare: a study in pigs. *Applied Animal Behaviour Science 40*: 82-83. https://doi.org/10.1016/0168-1591(94)90108-2

Izzo GN, Bashaw MJ and Campbell JB 2011 Enrichment and individual differences affect welfare indicators in squirrel monkeys *Saimiri sciureus*. *Journal of Comparative Psychology 125*: 347-352. https://doi.org/10.1037/a0024294

Keeling LJ, Rushen J and Duncan IJ 2011 Understanding animal welfare. In: Appleby MC, Mench JA, Olsson IAS and Hughes BO (eds) *Animal Welfare* pp 13-26. Cambridge University Press: Cambridge, UK. https://doi.org/10.1079/9781845936594.0013

Kistler C, Hegglin D, Wurbel H and Konig B 2009 Feeding enrichment in an opportunistic carnivore: The red fox. *Applied Animal Behaviour Science 116*: 260-265. https://doi.org/10.1016/j.applanim.2008.09.004

Lyon M and Robbins T 1975 The action of central nervous system stimulant drugs: A general theory concerning amphetamine effects. In: Essman W and Valzelli L (eds) *Current Developments in Psychopharmacology Vol 2* pp 81-163. Spectrum: New York, NY, USA

Major CA, Kelly BJ, Novak MA, Davenport MD, Stonemetz KM and Meyer JS 2009 The anxiogenic drug FG7142 increases self-injurious behavior in male rhesus monkeys *Macaca mulatta. Life Sciences 85*: 753-758. https://doi.org/10.1016/j.lfs.2009.10.003

Mason G and Mendl M 1993 Why is there no simple way of measuring animal welfare? *Animal Welfare 24*: 301-319

Mason GJ and Latham NR 2004 Can't stop won't stop: is stereotypy a reliable animal welfare indicator? *Animal Welfare 13*: S57-S70

Mellor DJ 2015 Positive animal welfare states and reference standards for welfare assessment. *New Zealand Veterinary Journal 63*: 17-23. https://doi.org/10.1080/00480169.2014.926802

Mellor DJ 2016 Updating animal welfare thinking: Moving beyond the 'Five Freedoms' towards 'a life worth living'. *Animals 6*: 21. https://doi.org/10.3390/ani6030021

Mendl M, Brooks J, Basse C, Burman O, Paul E, Blackwell E and Casey R 2010. Dogs showing separation-related behaviour exhibit a 'pessimistic' cognitive bias. *Current Biology 20*: R839-R840. https://doi.org/10.1016/j.cub.2010.08.030

Miller LJ, Mellen J, Greer T and Kuczaj SA 2011 The effects of education programmes on Atlantic bottlenose dolphin *Tursiops truncatus* behavior. *Animal Welfare 20*: 159-172

Miller LJ, Pisacane CB and Vicino GA 2016 Relationship between behavioural diversity and faecal glucocorticoid metabolites: a case study with cheetahs *Acinonyx jubatus. Animal Welfare 25*: 325-329. https://doi.org/10.7120/09627286.25.3.325

Napolitano F, Knierim U, Grass F and De Rosa G 2009 Positive indicators of cattle welfare and their applicability to on-farm protocols. *Italian Journal of Animal Science 8*: 355-365. https://doi.org/10.4081/ijas.2009.s1.355

Neal Webb SJ, Hau J and Schapiro SJ 2018 Captive chimpanzee (*Pan troglodytes*) behavior as a function of space per animal and enclosure type. *American Journal of Primatology 80*: e22749. https://doi.org/10.1002/ajp.22749

Pomerantz O, Terkel J, Suomi SJ and Paukner A 2012 Stereotypic head twirls, but not pacing, are related to a 'pessimistic'-like judgment bias among captive tufted capuchins (*Cebus apella*). *Animal Cognition 15*: 689-698. https://doi.org/10.1007/s10071-012-0497-7

Rabin LA 2003 Maintaining behavioural diversity in captivity for conservation: natural behaviour management. *Animal Welfare 12*: 85-94

Razal CB, Bryant J and Miller LJ 2017 Monitoring the behavioral and adrenal activity of giraffe *Giraffa camelopardalis* to assess welfare during seasonal housing changes. *Animal Behavior and Cognition 4*: 154-168. https://doi.org/10.12966/abc.03.05.2017

Rygula R, Pluta H and Popik P 2012 Laughing rats are optimistic. *PLOS One 7*: e51959. https://doi.org/10.1371/journal.pone.0051959

Shannon C and Weaver W 1949 *The Mathematical Study of Communication*. University of Illinois Press: Urbana, IL, USA.

Shannon CE 1948 A mathematical theory of communication. *Bell System Technical Journal 27*: 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shepherdson DJ, Carlstead K, Mellen JD and Seidensticker J 1993 The influence of food presentation on the behavior of small cats in confined environments. *Zoo Biology 12*: 203-216. https://doi.org/10.1002/zoo.1430120206

Stolba A, Baker N and Wood-Gush DGM 1983 The characterisation of stereotyped behaviour in stalled sows by informational redundancy. *Behaviour 87*: 157-182. https://doi.org/10.1163/156853983X00417

Tannenbaum J 1991 Ethics and animal welfare: The inextricable connection. *Journal of the American Veterinary Medical Association 198*: 1360-1376

Veasey JS, Waran NK and Young RJ 1996 On comparing the behaviour of zoo housed animals with wild conspecifics as a welfare indicator. *Animal Welfare 5*: 139-153

Vickery S and Mason G 2004 Stereotypic behavior in Asiatic black and Malayan sun bears. *Zoo Biology 23*: 409-430. https://doi.org/10.1002/zoo.20027

Wemelsfelder F, Haskell M, Mendl MT, Calvert S and Lawrence AB 2000 Diversity of behaviour during novel object tests is reduced in pigs housed in substrate-impoverished conditions. *Animal Behaviour 60*: 385-394. https://doi.org/10.1006/anbe.2000.1466

Wittig RM, Crockford C, Deschner T, Langergraber KE, Ziegler TE and Zuberbühler K 2014 Food sharing is linked to urinary oxytocin levels and bonding in related and unrelated wild chimpanzees. *Proceedings of the Royal Society, Series B 281*: 20133096. https://doi.org/10.1098/rspb.2013.3096