Behavioral persistence in captive bears: a response to Criswell and Galbreath

Sophie S. Vickery[1] and Georgia J. Mason[2]

*Animal Behaviour Research Group,
Department of Zoology, Oxford University,
South Parks Road, Oxford OX1 3PS, UK*

This paper responds to criticisms raised by Criswell and Galbreath (2005) regarding analyses we used in an earlier paper (Vickery and Mason 2003), and shows that their criticisms are unfounded. However, first, we present some background to the original work. Vickery and Mason (2003), written toward the end of a long-term study of captive bears, was timed to allow for presentation in 2002. One of us (SSV) was based at Banglamung Wildlife Breeding Station in Chonburi province, Thailand, a sanctuary which individually housed Asiatic black bears (*Ursus thibetanus*) and Malayan sun bears (*Helarctos malayanus*) in virtually identical conditions, providing a unique opportunity to investigate bears' behavioral responses to captivity. The behavior of 29 bears was observed over 2.5 years. This was the total number of individuals available for study, barring those that were sick or whose records were incomplete. Of these, 21 were put through learning tasks to test the hypothesis that their stereotypy stemmed from a general inability to switch behaviors appropriately. These learning tasks were time-consuming to run, and thus it was not possible to test all 29 bears. Twelve bears were tested in Year 1 (the data presented in Vickery and Mason [2003]), and a further 9 bears in Year 2. Both years' data are presented in Vickery and Mason (2005). In addition, the properties of all 29 bears' stereotypies were analyzed for effects of time in captivity, sex, and species (presented briefly in Vickery and Mason [2003], and more fully in Vickery and Mason [2004]).

It is well recognized that the stereotypies (repetitive, invariant and seemingly functionless forms of abnormal behavior [Ödberg 1978, Mason 1991]) of captive animals can have detrimental correlates. They have been associated with reduced behavioral diversity (Stolba et al. 1983, Shepherdson et al. 1993, Swaisgood et al. 2001), decreased responsiveness to environmental stimuli (Fentress 1977, Ormrod 1992, Langenhorst 1998, Clubb 2001), reduced exploratory behavior (Lien and Klopfer 1978), poor maternal competence (Mason et al. 1995), injury and reduced body condition (Sambraus 1985, Fraser and Broom 1990, Moon-Fanelli and Dodman 1998), and reduced reproductive success (Shepherdson 1994). Compared to their non-stereotypic counterparts, stereotypic animals may also have altered neurochemistry, for example showing alterations in dopamine (Dantzer 1986), serotonin (Vandebroek et al. 1995), and opioid (Zanella et al. 1996, Zanella and Mason 1997) functioning. Furthermore, recent work on bank voles (*Clethrionomys glareolus*), mice (*Mus musculus*), parrots (*Amazona amazonica*), and 2 species of songbird (*Parus caeruleus* and *P. palustris*) and rhesus monkeys (*Macaca mulatta*) has shown that stereotypic animals are more 'perseverative', i.e. more behaviorally persistent than are their less stereotypic conspecifics (Garner 1999, Garner and Mason 2002, Garner et al. 2003*a,b*, Lutz et al. 2004, Latham 2005). Our aim was to test for a similar correlation in bears, to investigate the generality of these findings. For species such as these, such an effect could also have conservation implications, potentially helping to explain the low reintroduction success currently displayed by captive-bred animals. In reintroduction programs, captive-bred animals generally fare far worse than translocated wild conspecifics (Griffith et al. 1989, Ginsberg 1994, Frantzen et al. 2001), and this certainly seems true for bears (Alt and Beecham 1984, Stiver et al. 1997, Eastridge and Clark 2001, Clark et al. 2002). For example, Stiver et al. (1997) reported that of 23 pen-reared American black bears (*Ursus americanus*) released in east Tennessee between 1982 and 1995, 6 were known to die, 5 caused nuisance problems, and 2 were killed illegally by humans.

The 2 main findings presented in Vickery and Mason (2003) were: (1) a positive correlation between an individual's stereotypy frequency and the number of trials for which it continued to respond without reward (a measure of behavioral persistence), and (2) a positive correlation between stereotypy frequency and time spent in captivity. Here, we divide Criswell and Galbreath's

---

[1]Present address: Animal Welfare Division, Department for Environment, Food and Rural, Affairs, 1A Page Street, London SW1P 4PQ, UK; sophie.vickery@defra.gsi.gov.uk
[2]Present address: Centre for the Study of Animal Welfare, Department of Animal and Poultry Science, University of Guelph, Ontario N1G 2W1, Canada; gmason@uoguelph.ca

criticisms of our work into 2 categories: those aimed at our general scientific approach, and those aimed at our specific statistical analysis.

Our aims and general scientific approach

Why did we do this work, and why did we use the subjects we did? One of Criswell and Galbreath's worries is that "It was never made clear ... why" 29 bears were used in one analysis and 12 in another. The implication is that we selected a non-random subset of animals from which it would clearly be illegitimate to extrapolate further. However, we did no such thing. This was an opportunistic study in which we simply used all the animals that were available to us and collected as much data as possible within logistic constraints. For example, we ran the learning experiments over 2 years because so few bears could be investigated simultaneously.

Second, Criswell and Galbreath claim that we "lumped all the bears available" into a single treatment group, thereby foregoing the opportunity to study a control group of bears. However, in fact, there was no appropriate control group, because all the bears that could be studied (i.e. those that were single-housed) were housed in near-identical cages and all except 2 exhibited stereotypy. It was for this reason that our study was correlational. Third, they claim that our "... inference would have been stronger had [we] used an experimental design in which those bears that had previously been removed from enriched enclosures to small cages were compared to bears lucky enough to remain behind in the enclosures." Here, we emphasize that we did not move the animals; rather our subjects were bears that had been housed in cages long before we arrived.

Finally, they incorrectly imply that we condemn reintroduction as a conservation tool. For example, they assert, "These findings lead Vickery and Mason, in turn, to question the potential success of reintroduction programs with bears held captive for substantial time periods". In fact, there was already ample evidence that reintroduction programs involving captive-bred bears, and indeed captive-bred animals of a range of species, are less successful than those involving their wild conspecifics. Our paper simply sought to raise and test ideas that might explain why this is so, with the aim of helping biologists improve reintroduction success in the future.

Statistical analyses

We categorize Criswell and Galbreath's statistical criticisms into those that are wrong and those that are right but unimportant. First, we deal with issues about which they are simply wrong. Criswell and Galbreath suggest that we would have obtained a better fit had we chosen ln S' (mean stereotypy levels across up to 6 observation periods between June 2000 and May 2002) in place of ln S (stereotypy levels measured immediately prior to task presentation). Although we appreciate the suggestion, this is irrelevant because these 2 stereotypy measures are far from interchangeable. Levels of stereotypy can vary with season, time in captivity, environmental change and other factors, and an individual's level of extinction responding might also vary over time (Garner 1999, Garner et al. 2003a). Thus, when investigating a relationship between stereotypy and extinction, it is sensible to measure these 2 variables as closely together as possible. The measure of stereotypy denoted S was used in our extinction analysis (Criswell and Galbreath [2005] Eq. 1) because this was taken immediately prior to task presentation. However, for the same reason — because an individual's stereotypy frequency might differ over the course of a year — when looking for a relationship between average stereotypy frequency and the number of years spent in captivity, it is most appropriate to assess stereotypy frequency on several occasions throughout the year and calculate an overall mean. Thus the measure used in our 'time in captivity' analysis (Criswell and Galbreath [2005] Eq. 2), S', represents a mean of stereotypy frequencies measured over up to 6 observation periods.

Second, they claim that it would be impossible to properly log-transform datasets containing zeros. In fact, this problem is commonly solved in practice by adding 1 to each value prior to transformation (Martin and Bateson 1993:151).

We now respond to statistical critiques that are technically correct but are ultimately unimportant for inference in this study. Criswell and Galbreath (2005) first claim that we reversed the "direction of causality". Our use of stereotypy as the dependent variable was appropriate because we were suggesting stereotypy to be a product of underlying changes that increase general behavioral persistence (see also Garner and Mason 2002). However, as Criswell and Galbreath (2005) point out themselves, "in a correlational study, that switch is not so serious." Our study was indeed exactly that—a correlational study, seeking to determine if 2 phenomena covaried. Furthermore, while they are correct that for the

first year's data alone ($n = 12$), if extinction is taken as the dependent variable the relationship with stereotypy becomes non-significant, if we repeat this analysis for the full dataset (n = 21; see Vickery and Mason 2005), swapping stereotypy and extinction in the model makes no difference to the significance of the effect (stereotypy as dependent variable—our preferred, for reasons of logic: $F = 5.49$; 1, 13 df; 1-tailed $P = 0.018$; extinction responding as dependent variable: $F = 5.49$; 1, 13 df; 1-tailed $P = 0.018$). Thus with a larger sample size the relationship holds regardless of whether stereotypy is deemed a dependent or independent variable.

Second, was our model "over-fitted"? Criswell and Galbreath (2005) claim that one of our models (their Eq. 1, which examines the relationship between stereotypy and behavioral persistence) included too many terms, leaving us with few remaining degrees of freedom. We counter that it would have simply been illogical to have ignored the variables incorporated here—they were included to control for potential confounds, i.e. other factors known to affect stereotypy or extinction learning (see below). For instance, although Criswell and Galbreath (2005) claim there were "complex interactions", in fact we included only one: species × sex, which they acknowledge has an important effect on stereotypy (see their Fig. 1). Many statisticians take the view that a model should partition out as much variance as possible (Grafen and Hails 2002, Kaps and Lamberson 2004), just as we did. Furthermore, if, for the sake of argument, we pare down the model using Minitab's best subsets regression command (which examines all possible regression models and identifies the model that gives the largest R^2 value; Minitab 12.0, Ryan and Joiner 2001), the best fitting model excludes the interaction term that Criswell and Galbreath are concerned about and leaves only time in captivity and extinction as the 2 most important explanatory variables. Importantly, in this model, the relationship between extinction and stereotypy holds ($F = 3.99$; 1, 9 df; 1-tailed $P = 0.039$). Finally, running our original (full) model on the complete dataset, with its greater degrees of freedom, gave us the same results (Vickery and Mason 2005).

Thirdly, Criswell and Galbreath (2005) state that we "failed, however, to report that their model as a whole fails to significantly explain variation in stereotypy frequency." While true, this is a red herring: this is simply irrelevant to the hypothesis under test. Whether a model is significant overall depends on whether non-significant variables within the model are omitted, or instead retained (as in our work). In a predictive model, non-significant variables are generally excluded so the final model retains only those variables that have predictive power—this will result in a final model that is, as a whole, significant (as was true for the reduced 'best subsets' model above: $P = 0.017$). However, if a variable is expected to have an effect on the dependent variable (for example, in this case, time in captivity was known to have an effect on stereotypy frequency), it should be included in the model even if it is non-significant, to act as a blocking factor (Grafen and Hails 2002). If non-significant variables are included in this way, the model as a whole may well be non-significant, but that does not invalidate any significant relationships that arise from it, which remain true and significant effects (Grafen and Hails 2002; F. Marriott, Oxford University Statistics Department, Oxford, UK, personal communication, 2003).

Fourth and finally, did we use models from entirely the wrong statistical family? Criswell and Galbreath criticize us for using parametric methods on bounded data: our stereotypy data were expressed as proportions of time, and thus logically could not be greater than 1 or less than 0. This makes them, strictly speaking, binomial. However, like many behavioral data, in practice they (or rather their residuals from the model) were not in fact significantly different from normal. (This is because, while theoretically bounded, in practice very few animals performed zero stereotypy, while — unsurprisingly—no animals spent 100% of their time stereotyping; thus even the raw data approximated normality). Furthermore, GLM's assumptions of homogenous variance were not violated, contrary to Criswell and Galbreath's claims. Data transformation is a commonly used and widely accepted method of correcting for non-normal distribution (Martin and Bateson 1993:150, Grafen and Hails 2002). Proportional data are commonly transformed using the angular (arcsine-square) transformation (Martin and Bateson 1993:150), but the appropriateness of any transformation always needs to be checked beforehand (Tabachnick and Fidell 1989, Grafen and Hails 2002). In the case of the Year 1 data reported in Vickery and Mason (2003), the angular transformation actually resulted in non-homogeneous variance, and a logarithmic transformation instead proved most appropriate. In contrast, no transformation was needed for the 'time in captivity' analysis. For both of these analyses, the residuals were normally distributed, thus confirming multivariate normality, that is, that each variable and all linear combinations of the variables were normally distributed (Tabachnick and Fidell 1989:70). Figures 1 and 2 present histograms and Anderson-Darling normality

(a)

(b)

Average: 0.0000000
StDev: 0.316752
N: 12

Anderson-Darling Normality Test
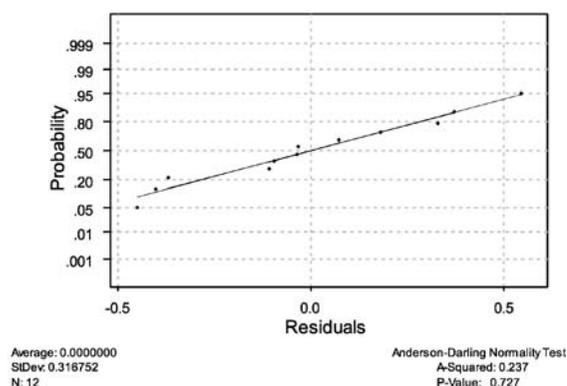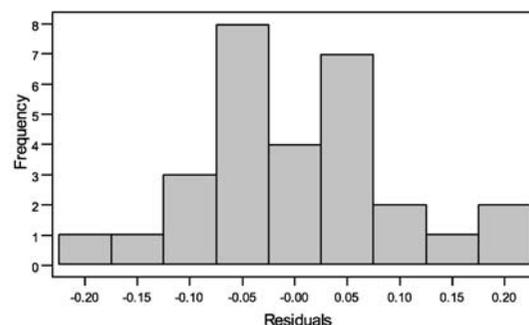A-Squared: 0.237
P-Value: 0.727
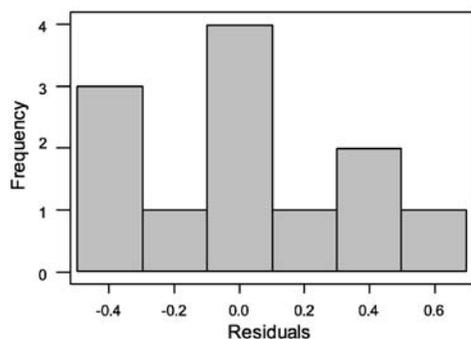
Fig. 1. Tests for normality of the residuals from the GLM analysis testing for a relationship between stereotypy frequency and responding in extinction (Equation 1; _n_ = 12): (a) histogram of residuals; (b) normality plot and Anderson-Darling Normality Test.

(a)

(b)

Average: 0.0000000
StDev: 0.0977666
N: 29

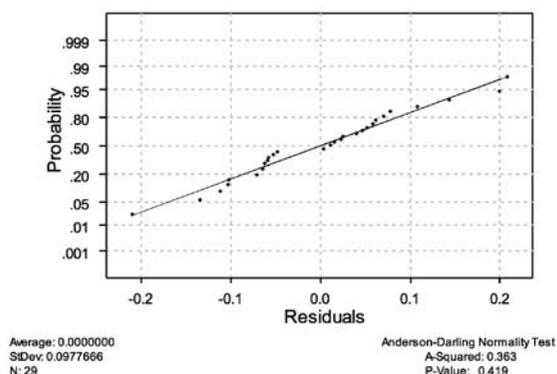Anderson-Darling Normality Test
A-Squared: 0.363
P-Value: 0.419

Fig. 2. Tests for normality of the residuals from the GLM analysis testing for a relationship between stereotypy frequency and time in captivity (Equation 2; _n_ = 29): (a) histogram of residuals; (b) normality plot and Anderson-Darling Normality Test.

plots and test statistics of the residuals for the extinction and time in captivity analyses respectively. Ryan-Joiner and Kolmogorov-Smirnov normality tests also verified that the residuals conformed to the normal distribution.

Given that none of the assumptions for using general linear models was actually breached, it was perfectly appropriate to use them (indeed, regressions and ANOVAs are commonly used on time budget data for just this reason; Martin and Bateson 1993), and it is unnecessary to insist on an alternative. Criswell and Galbreath's initial alternative resulted in a relationship between extinction and stereotypy that was even stronger than our original finding (Criswell and Galbreath 2005:Table 3), but which they claimed to be "biologically uninteresting." They then acknowledge problems with their own model's assumptions in the form of over-dispersion. This strikes us as unnecessary effort when our original model was perfectly legitimate.

Conclusion

The findings we presented in Vickery and Mason (2003) were neither controversial nor unprecedented.

We found a strong relationship between stereotypy and length of time in captivity (a factor correlating with bear age). This positive correlation was consistent with that found in other species, such as pigs (Cronin and Wiepkema 1984), mink (Mason 1993), mice (Würbel et al. 1996) and dogs (Siwak et al. 2001). We also found a positive relationship between stereotypy and behavioral persistence, just as we had predicted. Interestingly, to date this relationship has been found in all other species in which it has been investigated (Garner 1999; Garner and Mason 2002; Garner et al. 2003a,b; Latham 2005), raising pressing questions as to whether such effects reflect dysfunction (e.g. Vickery and Mason 2003). Furthermore, our concluding suggestions should not have been controversial. The reintroduction of captive-bred animals often fails, and it is important to understand why. We made no claims to have discovered the answer; we merely suggested a new, plausible, and testable hypothesis. We are thus glad that Criswell and Galbreath agree with our suggestion that bears housed in more natural, enriched environments might differ in their behavior from the individually caged bears that we tested (though their phrase 'learning adaptation' suggests some misunderstanding here; the effect is clearly not to do with learning *per se*). As our paper suggests, "Maintaining bears in large, naturalistic enclosures, in which they can perform a full range of natural behaviors and exert control over their environment, is likely to offer some protection against behavioral deficits." Testing this hypothesis in the future would be both fascinating and important.

Given all this, we found Criswell and Galbreath's criticisms of our paper to be a puzzling over-reaction. As we have shown, our analyses, findings and conclusions are not overturned by any of their criticisms. The manner in which these criticisms are presented also distressingly misses the key point: the reintroduction of captive bears is prone to failure and it is vital that biologists seek to understand why. Our real misdemeanors, such as they are, appear merely to be (1) publishing results from an *n* of just 12, instead of waiting until all our experimental data were processed (Vickery and Mason 2005), and (2) reminding readers of an uncomfortable truth: the very real current fallibility of reintroduction attempts.

Acknowledgments

Literature cited

ALT, G.L., AND J.J. BEECHAM. 1984. Reintroduction of orphaned black bear cubs into the wild. Wildlife Society Bulletin 12:169–174.

CLARK, J.D., D. HUBER, AND C. SERVHEEN. 2002. Bear reintroductions: lessons and challenges. Ursus 13:335–346.

CLUBB, R. 2001. The roles of foraging niche, rearing conditions and current husbandry on the development of stereotypies in carnivores. Dissertation, University of Oxford, Oxford, UK.

CRISWELL, A.R., AND GALBREATH, G.J. 2005. Behavioral persistence in captive bears: a critique. Ursus 16:268–273.

CRONIN, G.M., AND P.R. WIEPKEMA. 1984. An analysis of stereotyped behaviour in tethered sows. Annales de Recherches Veterinaires 15:263–270.

DANTZER, R. 1986. Behavioral, physiological and functional aspects of stereotyped behavior: review and a re-interpretation. Journal of Animal Science 62:1776–1786.

EASTRIDGE, R., AND J.D. CLARK. 2001. Evaluation of 2 soft-release techniques to reintroduce black bears. Wildlife Society Bulletin 29:1163–1174.

FENTRESS, J.C. 1977. The tonic hypothesis and the patterning of behavior. Annals of the New York Academy of Science 290:370–395.

FRANTZEN, M.A.J., J.W.H. FERGUSON, AND M.S. DE VILLIERS. 2001. The conservation role of captive African wild dogs (*Lycaon pictus*). Biological Conservation 100:253–260.

FRASER, A.F., AND D.M. BROOM. 1990. Farm animal behaviour and welfare. Bailliere Tindall, London, UK.

GARNER, J.P. 1999. The aetiology of stereotypy in caged animals. Dissertation, University of Oxford, Oxford, UK.

———, AND G.J. MASON. 2002. Evidence for a relationship between cage stereotypies and behavioural disinhibition in laboratory rodents. Behavioural Brain Research 136: 83–92.

———, ———, AND R. SMITH. 2003a. Stereotypic route-tracing in experimentally caged songbirds correlates with general behavioural disinhibition. Animal Behaviour 66: 711–727.

———, C.L. MEEHAN, AND J.A. MENCH. 2003b. Stereotypies in caged parrots, schizophrenia and autism: evidence for a common mechanism. Behavioural Brain Research 145(1–2):125–134.

GINSBERG, J.R. 1994. Captive breeding, reintroduction and the conservation of canids. Pages 365–383 *in* P.J.S. Olney, G.M. Mace, and A.T.C. Feistner, editors. Creative conservation: interactive management of wild and captive animals. Chapman and Hall, London, UK.

GRAFEN, A., AND R. HAILS. 2002. Modern statistics for the life sciences. Oxford University Press, New York, New York, USA.

GRIFFITH, B., J.M. SCOTT, J.W. CARPENTER, AND C. REED. 1989. Translocation as a species conservation tool: status and strategy. Science 245:477–480.

KAPS, M., AND W.R. LAMBERSON. 2004. Biostatistics for animal science. CABI Publishing, Wallingford, UK.

LANGENHORST, T. 1998. Effects of a behavioral enrichment program on the stereotypic behavior of brown bears (*Ursus arctos*). Der Zoologische Garten 68:341–354. (In German.)

LATHAM, N.R. 2005. Refining the role of stereotypic behaviour in the assessment of welfare: stress, general motor persistence and early environment in the development of abnormal behaviours. Dissertation, University of Oxford, Oxford, UK.

LIEN, J., AND F.D. KLOPFER. 1978. Some relations between stereotyped suckling in piglets and exploratory behaviour and discrimination reversal learning in adult swine. Applied Animal Behaviour Science 4:223–233.

LUTZ, C., S. TIEFENBACHER, J. MEYER, AND M. NOVAK. 2004. Extinction deficits in male rhesus macaques with a history of self-injurious behaviour. American Journal of Primatology 63:41–48.

MARTIN, P., AND P.P.G. BATESON. 1993. Measuring behaviour: an introductory guide. Second edition. Cambridge University Press, Cambridge, UK.

MASON, G.J. 1991. Stereotypies: a critical review. Animal Behaviour 41:1015–1037.

———. 1993. Age and context affect the stereotypies of caged mink. Behaviour 127:191–229.

———, A, LEIPOLDT, AND G. DE JONGE. 1995. Why do female mink with high stereotypy levels have slow-growing offspring? Pages 133–143 *in* S.M. Rutter, J. Rushen, H.D. Randle, and J.C. Eddison, editors. Proceedings of the 29th International Congress of the International Society for Applied Ethology. Universities Federation for Animal Welfare, Wheathampstead, Herts, UK.

MOON-FANELLI, A.A., AND N.H. DODMAN. 1998. Description and development of compulsive tail chasing in terriers and response to clomipramine treatment. Journal of the American Veterinary Medical Association 212:1252–1257.

ÖDBERG, F.O. 1978. Abnormal behaviours: (stereotypies). Pages 475–480 *in* J. Garsi, editor. Proceedings of the First World Congress on Ethology Applied to Zootechnics. Industrias Graficas, Madrid, Spain.

ORMROD, S.A.A. 1992. A review of captive polar bears in Great Britain and Ireland. Report commissioned by the Born Free Foundation, Horsham, West Sussex, UK.

RYAN, B.F., AND B.L. JOINER. 2001. Minitab handbook. Fourth edition. Duxbury, Pacific Grove, California, USA.

SAMBRAUS, H.H. 1985. Stereotypies. Pages 431–441 *in* A.F. Fraser, editor. Ethology of farm animals. Elsevier, Amsterdam, The Netherlands.

SHEPHERDSON, D.J., K. CARLSTEAD, J.D. MELLEN, AND J. SEIDENSTICKER. 1993. The influence of food presentation on the behavior of small cats in confined environments. Zoo Biology 12:203–216.

———. 1994. The role of environmental enrichment in the captive breeding and reintroduction of endangered spe-cies. Pages 167–177 *in* P.J.S. Olney, G.M. Mace and A.T.C. Feistner, editors. Creative conservation: interactive management of wild and captive animals. Chapman Hill, London, UK.

SIWAK C.T., TAPP P.D., AND MILGRAM N.W. 2001. Effect of age and level of cognitive function on spontaneous and exploratory behaviors in the beagle dog. Learning and Memory 8:317–325.

STIVER, W.H., M.R. PELTON, AND C.D. SCOTT. 1997. Use of pen-reared black bears for augmentation or reintroductions. International Conference on Bear Research and Management 9(2):145–150.

STOLBA, A., N. BAKER, AND D.G.M. WOOD-GUSH. 1983. The characterization of stereotyped behaviour in stalled sows by information redundancy. Behaviour 87:157–182.

SWAISGOOD, R.R., A.M. WHITE, X. ZHOU, H. ZHANG, G. ZHANG, R. WEI, V.J. HARE, E.M. TEPPER, AND D.G. LINDBURG. 2001. A quantitative assessment of the efficacy of an envi-ronmental enrichment programme for giant pandas. Animal Behaviour 61:447–457.

TABACHNICK, B.G., AND L.S. FIDELL. 1989. Using multivariate statistics. Second edition. HarperCollins, New York, New York, USA.

VANDEBROEK, I., F.O. ÖDBERG, AND J. CAEMAERT. 1995. Microdialysis study of the caudate nucleus of stereotyp-ing and non-stereotyping bank voles. Pages 245–246 *in* S.M. Rutter, J. Rushen, H.D. Randle, and J.C. Eddison, editors. Proceedings of the 29th International Congress of the International Society for Applied Ethology. Universities Federation for Animal Welfare, Wheathampstead, Herts, UK.

VICKERY, S.S., AND G.J. MASON. 2003. Behavioral persistence in captive bears: implications for reintroduction. Ursus 14:35–43.

VICKERY, S., AND G. MASON. 2004. Stereotypic behavior in Asiatic black and Malayan sun bears. Zoo Biology 23:409–430.

———, AND ———. 2005. Stereotypy and perseverative responding in caged bears: further data and analyses. Applied Animal Behaviour Science 91(3–4):247–260.

WÜRBEL, H., M. STAUFFACHER, AND D. VON HOLST. 1996. Ste-reotypies in laboratory mice — Quantitative and qualitative description of the ontogeny of 'wire-gnawing' and 'jumping' in Zur:ICR and Zur:ICR nu. Ethology 102:371–385.

ZANELLA, A.J., D.M. BROOM, J.C. HUNTER, AND M.T. MENDL. 1996. Brain opioid receptors in relation to stereotypies, inactivity and housing in sows. Physiology and Behavior 59:769–775.

———, AND G.M. MASON. 1997. Neurophysiological corre-lates associated with behavioral traits in mink. Third North American Regional Meeting of the International Society for Applied Ethology, www.applied-ethology.org.