Zoo Biology 25:317–337 (2006)

Research Article

Meta-Analytic Review of the Effects of Enrichment on Stereotypic Behavior in Zoo Mammals

Amanda Shyne*

Department of Psychology, Northeastern University, Boston, Massachusetts

This meta-analysis reports the effect enrichment has on the occurrence of stereotypic behavior exhibited by captive zoo mammals. The analysis also identifies which types of enrichment are most effective, which groups of animals benefit the most, and which types of stereotypes are most affected by environmental enrichment. The analysis included 54 studies that yielded 63 effect size statistics. Fifty-seven of sixty-three effect sizes went in the predicted direction (90%), with the animals participating in less stereotypic behavior during the enrichment condition than in the baseline condition. The mean effect size (correlation coefficient r) was 0.46. The combined P-value using both fixed and random effects methods was reveled to be <0.0000001. A file drawer N-value was calculated to identify the number of unretrieved studies (with a combined effect size of zero) that would be needed to nullify the results of this analysis. The file drawer N-value was 1,726, suggesting that it is highly unlikely that the significant results reported in this analysis are nullified by studies that remain in file drawers. Based on these results it was concluded that enrichment substantially reduces the frequency of stereotypic behavior exhibited by mammals living in zoo environments. Zoo Biol 25:317–337, 2006. © 2006 Wiley-Liss, Inc.

Keywords: environmental enrichment; animals; stereotypes; captivity

INTRODUCTION

Stereotypic behavior is invariant, repetitive, and has no obvious goal or function [Keiper, 1969; Rushen, 1984; Mason, 1991]. Common types of stereotypic patterns in zoo animals include pacing, head rolling, excessive licking, hair or feather

*Correspondence to: Amanda Shyne, PhD, Northeastern University, Psychology Department, 125 Nightingale Hall, Boston, MA 02115. E-mail: amanda.shyne@gmail.com

Received 23 September 2004; Accepted 16 December 2005

DOI 10.1002/zoo.20091
Published online 19 April 2006 in Wiley InterScience (www.interscience.wiley.com).

plucking, and pattern swimming. Stereotypes often arise when a captive animal has prolonged exposure to an ecologically relevant problem that it is incapable of solving within its enclosure. Such ecologic problems include finding a mate, hunting for food, and escaping from human contact [Shepherdson, 1989]. The frustration caused by the animal's inability to carry out certain goal-directed behaviors often leads to stereotypic patterns. For example, laying hens that are not provided with appropriate nesting sites will engage in stereotypic pacing [Duncan, 1970].

Stereotypes may also develop if the captive animals are forced to live in environments with persistent unavoidable stress and fear. Examples include chaffinches who show an increased amount of stereotypic preening when their cages are located near stuffed predator owls [Rowell, 1961] and a female sloth bear that exhibited stereotyped vomiting and swaying when forced into contact with an aggressive male [Meyer-Holzapfel, 1968]. Finally, stereotypes have been associated with barren enclosures that do not provide stimulation. In sub-optimal barren environments stereotypes are more evident than in enclosures that provide species-appropriate environmental stimuli [Mason, 1991; Mason and Latham, 2004]. For example, natural foraging behavior in pigs consists of a complex set of behaviors but these same foraging behaviors are expressed in fixed and stereotyped patterns when pigs are living in barren enclosures. The foraging behaviors (rooting and digging) are simplified and directed toward static features of the animals' environment such as fence bars [Rushen et al., 1993]. In captivity, ecologically motivating factors can stimulate the animal to engage in an initial behavior that is not rewarded in captivity, and over time this behavior can become repetitive and decreasingly reliant on the environmental cues that in the natural world regulate the behavior [Mason, 1993].

Stereotypic patterns of behavior can also restrict animals from interacting with their environment in a normal species typical fashion. Animals that participate in pervasive stereotyped fixations sometimes show an overall reduction in behavioral diversity [Dantzer, 1986; Garner et al., 2003], abnormal disintegration of species typical behavior patterns, and the inability to interact appropriately with novel stimuli [Wemelsfelder, 1993]. Therefore, stereotypic behavior provides direct evidence for impoverished welfare and suffering [Broom, 1983].

In conjunction with the more widely recognized stereotypic fixations (pacing, head rolling, etc.), abnormal behaviors such as regurgitation/reingestion and coprophagy (feeding on excrement) are commonly observed in zoo primates. These behaviors are similar to stereotypes because they often become ritualized and repetitive; because of this, some authors have referred to these behaviors as forms of stereotype [Hinde, 1970; Loeffler, 1982]. More current research, however, refers to them as abnormal behaviors that are observed infrequently in the wild [Akers and Schildkraut, 1985; Gould and Bres, 1986; Ruempler, 1992]. Regardless of how these (abnormal) behaviors are classified, their occurrence is a problem as they are extremely pervasive in captive primates (especially gorillas), and due to their unpleasant nature, the behaviors often disgust the zoo public [Akers and Schildkraut, 1985].

In response to concerns about animals' welfare and the educational experience of the zoo, many zoos are implementing enrichment programs to encourage natural behavior patterns that approximate behaviors found in the wild [Chamove, 1989]. In these cases, the role of environmental enrichment is to provide the captive animal with the stimuli needed to participate in species-typical behavior and to increase the amount of time the animals spend in goal-directed behaviors. Concurrently, enrichment methods should reduce the amount of time spent in both stereotypic behaviors and abnormal behaviors [Wemelsfelder, 1993].

The scientific study of enrichment has used various behavioral and biologic methods to evaluate the effect of enrichment interventions. In a 1995 review, Newberry [1995] defined enrichment as an improvement in the biologic health of the animal resulting from modifications to the enclosure. The animal's biologic health was measured by its inclusive fitness and reproductive success. Chamove [1989] defined enrichment as methods that alter the captive animal's behavior so that the behaviors observed are similar to the behaviors expressed by the animal's wild conspecifics. Lastly, the definition used by Mellen and McPhee [2001] combined both biologic and behavioral measures. In the current meta-analytic review, environmental enrichment is evaluated using a behavioral marker, the frequency of stereotypic behavior. Chamove [1989] states that the frequency with which animals engage in stereotypic behavior is a sensitive measure of the effectiveness of an enrichment condition.

The previous review articles on enrichment [Chamove, 1989; Newberry, 1995; Mellen and MacPhee, 2001] discuss operational definitions for enrichment, the different types of enrichment, and the methods to use when implementing an enrichment study. These review articles have been improved on by the work of Swaisgood and Shepherdson [2005]. In their article, the authors reviewed the literature and recorded whether or not each enrichment study included had a statistically significant effect. Combining the data in this way is called "vote counting" [Lipsey and Wilson, 2001]. The systematic vote counting methods used by the authors provide a more quantitative assessment of the literature than had been done previously. Unfortunately, vote counting is a review procedure that is often misleading. Statistical significance is dependent on both the magnitude of the effect and the sampling error around the estimate, which is a function of sample size [Lipsey and Wilson, 2001]. Swaisgood and Shepherdson [2005] acknowledged this shortcoming and stated that because of the consistently small sample sizes included in their review the method was valid. But even if significance levels are not confounded by sample size when the sample sizes are small, as they typically are in this literature, the vote-counting method can lead to an under appreciation of the magnitude of the effect. Statistical significance in the primary literature is dependent on both the magnitude of the effect and the sampling error, which is dependent on the sample size of the individual study. A meta-analysis compiles effect sizes, which are independent of sample size, and therefore, are not biased toward larger studies. Sophisticated meta-analytic techniques encode the magnitude as well as the direction of each statistical relationship within the literature. The effect sizes calculated in a meta-analysis provide a variable sensitive to the fluctuating strengths of the different studies analyzed [Lipsey and Wilson, 2001].

PRESENT META-ANALYTIC REVIEW

The purpose of the current comprehensive meta-analytic review is to systematically examine the influence of various forms of environmental enrichment on negative stereotypic behaviors expressed by animals living in zoos. Meta-analytic techniques were used due to the fact that they are more systematic and quantitative than traditional reviews. By encoding the magnitude of each relevant statistical relationship found in every published article and then compiling the effect size statistics, the meta-analytic review provides a sensitive tool for finding the strength of an effect and can conclusively demonstrate statistical significance.

In addition to being a quantitative assessment of the literature, meta-analyses are also sensitive to finding effects that are obscure in the individual studies themselves. The large numbers of studies (and their corresponding effect sizes) used in meta-analytic procedures provide the data needed to perform contrasts on possible moderating variables. Moderating variables are variables that are associated with effect magnitude [Hall et al., 1994]. For example, in enrichment research, the taxonomic group of the animals studied may affect the success of the enrichment manipulation but few primary literature articles have the data needed to analyze this important question. Other possible moderating variables include the form of the enrichment and the type of the stereotype exhibited. Unfortunately due to the qualitative nature of traditional reviews, moderating variables such as these are impossible to analyze.

MATERIALS AND METHODS

Literature Searches

Extensive searches of the psychologic and biologic literature were carried out to locate articles on the relationship between enrichment and stereotypes. First, two online computer reference databases were searched: PsycINFO and Biosis. The terms used in the search were enrichment, stereotypes, zoo animals, and animal welfare. These terms were searched separately and crossed with each other. Second, the reference section of reviews on both environmental enrichment and stereotypic behavior were scrutinized. Last, book chapters on related topics were combed for relevant references for the current review.

Inclusion Criteria

The criteria for including studies in the sample were: 1) the dependent variable was time spent performing stereotypic behavior (pacing, head rolling, etc.) or abnormal behavior (regurgitation/reingestion, coprophagy, and drinking urine); 2) the results were based on any species of captive zoo animal or animal in a zoologic park (including animals typically thought of as domestic stock animals or pets if they were kept in the zoo environment); 3) the study was published and written in English; 4) a form of enrichment was introduced into the animal's already existing exhibit (including behavioral training in the exhibit); and 5) the study compared a baseline period to an enrichment period. All studies that met these criteria were included, regardless of sample size; i.e., there was no minimum sample size required (even $N = 1$).

Exclusion Criteria

Studies were omitted if enrichment was in the form of 1) removing negative stimuli from the exhibit; 2) zoo visitors; 3) additional intraspecific or interspecific animals; or 4) visual presence of predator or prey.

Studies in Sample

The resulting sample of 54 studies yielded 63 effect size statistics. Of these 63 effect sizes, 53 were independent (i.e., were based on different groups of animals).

Variables Coded From Each Study

The following information was recorded for each report: 1) species; 2) publication year; 3) number of animals; 4) the animals' gender; 5) if the animals had reached sexual maturity; 6) composition of the group (how the animals were housed, either in same sex groups, mixed sex groups, or isolation); 7) number of enrichment interventions used in study; 8) type of enrichment intervention (food puzzle includes scatter feeding, artificial prey, physical toy, additional furniture, training, intact carcass, other); 9) methodology; 10) location in which the animals were housed (name of the zoo or zoologic park); 11) type of stereotypic behavior pattern (pacing, swaying, self-mutilation, oral fixation, coprophagy, head rolling, regurgitation and reingestion, other); and 12) whether or not the enrichment encouraged the animal to participate in more natural behavior. The enrichment type, animal species, and form of stereotype were extracted from each study to analyze the effect of these moderating variables on the effect size.

Initially this review included all enrichment studies and coded all the different dependent variables used by the individual researchers. However, when all possible dependent variables were included, four problems became relevant. First, many of the studies violated assumptions of independence. For observations to be independent of one another, a single observation must have no influence on any other observation. For example, studies that treat data collected from one tiger over multiple days as if it were multiple tigers violate independence assumptions, due to the likelihood that the observations have a consistent and predictable relationship to one another. Second, there was a large number of possible dependent variables and many researchers looked at a substantial number but only reported the significant findings. Third, there was very little consensus on how the dependent variables would be affected positively by the enrichment intervention. For example, some studies claimed that increased activity was beneficial for the animals (and in the correct direction) and some studies claimed that a decrease in activity was beneficial (and in the correct direction). Although these assumptions were probably correct based on the natural history of the species of animal being studied, it made systematic analysis difficult. Last, there was much ambiguity in the terminology used to describe simple behaviors. For example, is walking the same as locomoting (or swimming)? The current meta-analytic review resolved these problems by concentrating only on the effect various types of enrichment had on the frequency of stereotypic and abnormal behaviors expressed in the captive zoo population. These two dependent variables, frequency of stereotype and frequency of abnormal behavior, were always viewed as negative. Therefore, researchers consistently predicted that the enrichment manipulations used in their studies would improve welfare as shown by a reduction in one of these types of behavior [Kastelein and Wiepkema, 1989; Baxter and Plowman, 2001].

Study Characteristics

Of 63 effect sizes calculated in this analysis none came from studies using non-mammalian zoo animals. Although no specific criteria excluded non-mammals, the few studies using non-mammals were eliminated because they did not meet the inclusion criteria. One of the most common reasons the studies using non-mammalian animals were excluded was because they did not use stereotypic/

abnormal behaviors as their dependent variable [Cardiff, 1996; Burghardt et al., 1996; Bauck, 1998; Anderson and Wood, 2001].

Dates of publication ranged from 1978–2003. Although the literature search and inclusion criteria had no date requirement, all of the systematic research in the field was quite recent, the earliest study being published (with quantitative data) in 1978. Fifty-three percent of the studies were published from 1999–2003, and 84% of the studies were published from 1990–2003.

Meta-Analytic Techniques Used

In this review, the correlation coefficient (r) was used as the effect size estimate. (Effect size values are not affected by the sample size of the original studies.) The correlation coefficient was chosen because it is easily computed from χ^2, F, t, d, or from the standard normal deviate Z and because it is easily interpretable [Rosenthal, 1991]. The dependent variable in this analysis was the proportion of stereotypic behavior. Many of the studies used in this meta-analysis only reported raw data. In these cases, a matched sample t-test between pre- and post-intervention behavior was carried out [Lipsey and Wilson, 2001]. Then t or whatever test statistic was given in the publication was converted (using standard formulas) into the correlation coefficient r. The mean effect sizes were computed using the correlation coefficient (r) and Fisher's r-to-z transformation (rz). This rz conversion was carried out to normalize the data and to make the intervals equal. Fisher's rz gives heavier weight to effect sizes that are further from zero [Rosenthal and Rosnow, 1991]. In addition, weighted (by sample size) mean effect sizes were computed using both effect size r's and Fisher's rz's (that are then converted back to r). When computing the weighted mean, the effect size was multiplied by the total number of subjects in the study, then those numbers were summed and divided by the total number of subjects across studies. Often meta-analysts use N-3 when computing the weighted mean [Rosenthal, 1991], but with the small sample sizes used in this literature that statistic could not be calculated.

Indices of effect size r and Fisher's rz were given positive signs when the enrichment reduced the amount of stereotypic behavior. Negative values were assigned when the subjects increased stereotypic behaviors during the enrichment condition.

All the results from each study were combined and a P-value was assigned to the compilation. This was done using a fixed effects method and a random effects method. The fixed effects method used the following formula to find P, $Z = \sigma Z/\sqrt{N}$ [Rosenthal, 1991], after finding the combined Z a table was referenced to find the corresponding P-value. This fixed effects method is relevant to the methodologies used for the particular set of studies under analysis, whereas the random effects method allows for inferences to be made to new studies with different methodologic features [Lipsey and Wilson, 2001]. The fixed effect method is simply a one-sample t-test comparing the averaged r values to zero [Rosenthal, 1995]. Whenever possible it is desirable to calculate both fixed effects and random effects.

In addition, a file drawer N was calculated. This statistic is used because of the tendency for journals to publish studies with significant results. Presumably, similar studies, which do not reach significance, remain in file drawers or are unretrieved for other reasons. The file drawer N gives the number of studies (that have an average effect size of 0) that would be needed to nullify the results of the meta-analysis, to specifically bring the combined P up to a non-significant level [Rosenthal, 1991].

With the amount of variation contained in this review for both the subjects (different species of animal, different ontogeny, etc.) and the types of enrichment (food puzzles, artificial prey, etc.), it is unclear if the effect sizes calculated in this meta-analysis came from the same population. Therefore a homogeneity test was calculated to see if the dispersion of the effect sizes around the mean was due to sampling error alone or if the data came from separate populations [Rosenthal, 1991; Lipsey and Wilson, 2001]. Typically, when performing a meta-analysis the equation is $\sigma(\text{n-3})(rz-\text{mean } rz)^2 = X^2$ with the $df = \text{K-1}$ (K was defined as the number of studies), but in this particular meta-analysis with the sample sizes averaging five subjects, n was substituted for n-3.

Finally, a number of contrasts were computed to see if there was statistical significance in the relationship between specific moderator variables and their obtained effect sizes. Often researchers believe that the test of heterogeneity must be found significant before contrasts can be computed. This is not the case for planned contrasts [Hall and Rosenthal, 1995; Rosenthal, 1995] as in the present case where a priori three coded variables were analyzed to observe their effect on the size of the Fisher *rz*. These three variables were the type of enrichment intervention, the taxonomic group of the subjects, and the type of stereotypic behavior exhibited by the animals. Three contrasts were carried out on the variable "enrichment intervention" to see if different forms of enrichment significantly affected the size of the Fisher *rz*. The three forms of enrichment that were analyzed to quantify their influence on the effect size were scent enrichment, food puzzles, and additional furniture. Because of the variability between these different enrichment types, it was predicted that the effects of each form of enrichment would differ in their ability to reduce stereotypic behavior. Therefore, each of these three forms of enrichment was contrasted, individually, against all the other studies used in the meta-analysis. For example, when carrying out the contrast on food puzzle enrichment the 39 studies that used various forms of food enrichment were contrasted against all the other studies that used a different type of enrichment manipulation. Thirty-nine of sixty-three effect size statistics used food puzzles as the form of environmental enrichment. Eight effect size statistics came from studies that used additional furniture as the enrichment manipulation and four effect sizes came from studies that used scent enrichment. The other forms of environmental enrichment included computer games, physical toys, wood chip litter, and artificial prey.

The second moderator variable analyzed was the taxonomic group of the animal. This was investigated to uncover possible differences that may have existed with regard to how different species of animal react to enrichment. Because there were not enough studies to perform meaningful contrasts on individual species, however, contrasts were carried out on only three groups of animals: terrestrial carnivores, marine mammals, and non-human primates. These groupings were chosen, primarily, because these types of mammals were well represented in the studies used in this analysis. Each of these three groups was contrasted, individually, against all the other studies in the analysis. The meta-analysis included 37 effect size statistics using carnivores as subjects, 10 using marine mammals, and 13 using primates.

The third variable predicted to affect the size of the Fisher *rz* was the type of stereotypic behavior carried out by the zoo animals. The two forms of stereotype analyzed were pacing (including pattern swimming) and abnormal behavior. These behaviors were analyzed because pacing was the most common stereotype reported and abnormal behavior is not always considered a stereotypic behavior. Pacing/pattern swimming had 41 effect size statistics and abnormal behavior had 10 effect size statistics.

Unknown Effect Size Statistics

Of the 63 effect sizes used in this analysis, 17 were given a value of zero. This occurred when the information necessary for finding an effect size statistic was missing from the original publication. Of these unknown effect size indices, 12 went in the predicted direction, with the animals in the baseline condition expressing more stereotypic behavior than in the enrichment condition. Two of the unknown effect sizes went in the reverse direction, with the animals participating in more stereotypy in the enrichment condition than in the baseline condition. And, last, three of the effect sizes were completely unknown. To remain conservative and decrease the possibility of type one error, all unknown effect sizes were assigned a value of zero. Many of the analyses carried out for this meta-analysis were done twice: once on the data when the effect size statistics assigned a zero value were included and once when the unknown effect size indices were excluded.

RESULTS

Measures of Central Tendency

Fifty-seven of sixty-three effect sizes went in the predicted direction (90%), with the animals participating in less stereotypic behavior during the enrichment condition than in the baseline condition (Table 1). The mean unweighted effect size for the correlation coefficient r was 0.46 when the assigned zero values were included

TABLE 1. Enrichment and stereotypic behavior effect size statistics and selected sample characteristics

Study	Species	Enrichment type	Stereotype	Effect size r	Fisher's rz
Wozniak, 1999	Okapi	Food puzzle	Pacing	0.73	0.93
Ames, 1999	Polar bear	Additional furniture	Pacing stereotypic swimming	0.71	0.89
Tepper et al., 1999	Giant panda	Food puzzles Physical toys	Pacing	0.71	0.89
Williams et al., 1999	Asiatic lion Sumatran tiger	Scent	Pacing	0.687	0.848
Bowkett et al., 1999	Tufted capuchin monkey	Food puzzle	Pacing	0.92	1.53
Angele et al., 1999	Asian elephant	Food puzzle Physical toys	Swaying	0.71	0.89
Carlstead and Seidensticker, 1991	American black bear	Odor	Pacing	0.33	0.34
	American black bear	Food puzzle	Pacing	0.64	0.76

TABLE 1. *Continued*

Study	Species	Enrichment type	Stereotype	Effect size r	Fisher's rz
Kolter and Zander, 1995	Polar bear	Food puzzle	Pacing	0.93	1.66
Brown and Gold, 1995	Lowland gorilla	Additional furniture	Coprophagy	0.45	0.49
	Lowland gorilla	Additional furniture	Regurgitation and reingestion	0.48	0.52
	Lowland gorilla	Additional furniture	Holding ears	0.35	0.37
Hare, 1995	Alaska brown bear	Food puzzle	Pacing	0.90	1.47
Gould and Bres, 1986	Gorilla	Food puzzle	Regurgitation and reingestion	0.81	1.13
Schiess-Meier, 1997	Snow leopard	Physical toy	Pacing	0.66	0.79
Hartmann, 1997	European wildcat	Food puzzle	Pacing	+0	
Yanofsky and Markowitz, 1978	Mandrill	Computer game	Pacing	0.97	2.09
Chamove et al., 1982	Seven primate species	Additional furniture	Abnormal behavior	+0	
Stoinski et al., 2000	African elephant	Food puzzle	Swaying	–0	
Morimura and Ueno, 1999	Brown bear	Food puzzle	Context free behavior	0.74	0.95
	Asian elephant	Food puzzle	Head swinging	0.98	2.30
Swaisgood et al., 2001	Giant panda	Food puzzle Physical toys	Self mutilation Oral stereotypes Head rolling	0.173	0.172
Kastelein and Wiepkema, 1988	Steller sea lion	Training	Circle swimming	0.80	1.099
Kastelein and Wiepkema, 1989	Pacific walrus	Food puzzle	Circle swimming	0.80	1.10
Grindrod and Cleaver, 2001	Common seal	Food puzzle Artificial prey Physical toys	Circle swimming	0.76	1.0
Hunter et al., 2002	Harbor seal, gray seal	Food puzzle, Artificial prey, Physical toys	Pattern swimming	0.51	0.56
Ross, 2002	Asian small-clawed otter	Food puzzle	Hair plucking	0.86	1.29
Baxter and Plowman, 2001	Giraffe	Food puzzle	Oral stereotype	0.45	0.49
Jenny and Schmid, 2002	Amur tiger	Food puzzle	Pacing	0.84	1.22
Lyons and Young, 1997	Seven species of cat	Food puzzle	Pacing	+0	
Carlstead, 1991	Fennec fox	Additional furniture	Pacing	–0.82	–1.16
		Artificial prey	Pacing	0.84	1.22
		Food puzzle	Pacing	0.59	0.68

TABLE 1. *Continued*

Study	Species	Enrichment type	Stereotype	Effect size *r*	Fisher's *rz*
Shepherdson et al., 1989	Slender-tailed meerkat	Artificial prey	Pacing	+0	
Wechsler, 1992	Polar bear	Odors	Pacing	−0.04	−0.04
Renner and Lussier, 2002	Spectacled bear	Food puzzle Additional furniture	Pacing	0.71	0.89
Paquette and Prescott, 1988	Chimpanzee	Physical toy	Abnormal behavior Coprophagy Licking walls Drinking urine	0.51	0.56
Altman, 1999	Polar bear, sloth bear, spectacled bear	Physical toy	Pacing	0.53	0.59
Fischbacher and Schmid, 1999	Spectacled bear	Food puzzle	Head tossing	?0	
Forthman et al., 1992	Polar bear, Kodiak bear, Asiatic black bear	Food puzzle	Pacing Stereotyped swimming	0.83	1.19
Carlstead et al., 1991	Sloth bear, American black bear, brown bear	Food puzzle	Pacing	0	0
Markowitz and LaForse, 1987	Serval	Artificial prey	Pacing	+0	
McPhee, 2002	Large felid	Intact carcass	Pacing	+0	
Shepherdson et al., 1993	Leopard cat	Food puzzle	Pacing	0.89	1.42
Markowitz et al., 1995	African leopard	Artificial prey	Pacing	0.15	0.15
Carlstead et al., 1993	Leopard cat	Additional furniture	Pacing	0.83	1.19
Bashaw et al., 2003	Sumatran tiger	Food puzzle Live fish	Pacing	0.35	0.37
	African lion, Sumatran tiger	Bones	Pacing	0.59	0.68
Wood, 1998	Chimpanzee	Novel enrichment	Coprophagy Fecal art Self-plucking of body hair Stereotypic movements	0.92	1.59
Rooney and Sleeman, 1998	Gorilla	Food puzzle Physical toys	Regurgitation and reingestion Coprophagy	0	0
Cohen and Moore, 2001	Polar bear	Odor	Pacing	−0	−0
Plowman and Knowles, 2001	Tiger	Food puzzle Odors Physical toys	Pacing	+0	+0
Small, 2001	Sun bear	Food puzzle	Pacing	0.57	0.65

TABLE 1. *Continued*

Study	Species	Enrichment type	Stereotype	Effect size r	Fisher's rz
Vivian et al., 2001	Chimpanzee	Food puzzle Additional furniture Physical toys	Abnormal behavior	+0	
Cheng, 2001	Sun bear	Food puzzle	Pacing	0.98	2.30
	Sun bear	Food puzzle	Tongue protruding	0.58	0.66
	Sun bear	Food puzzle	Paw licking	0.58	0.66
Koene et al., 2001	Felid	Intact carcass	Stereotypic behavior	+0	
Veninga and Lemon, 2001	African wild dog	Intact carcass	Pacing	+0	
Wiard, 1992	Gorilla	Food puzzle	Regurgitation and reingestion	+0	
Gatcliffe et al., 2002	Giraffe	Food puzzle	Oral stereotypes	0.52	0.58
Wehnelt and Hudson, 2002	Felid	Food puzzle	Pacing	0.64	0.76
Holden, 2003	Harbor seal	Food puzzle	Circle swimming	+0	

+0, unknown effect sizes that went in predicted direction; 0, unknown effect sizes that went in reverse direction; ?0, unknown effect sizes.

and when the zero values were excluded the *r* was 0.62. The median *r* was 0.55 with all 63 effect-sizes included. When the Fisher *rz* values were transformed back into the correlation coefficient *r* the value for all effect sizes combined was 0.57 and when the assigned zero values were excluded it was 0.71.

The mean weighted (by sample size) effect size *r* for all 64 effect-size indices was 0.36. The mean weighted effect size *r* was 0.60 when the unknown values assigned a zero score were excluded. When Fisher *rz* values were transformed back into the correlation coefficient *r* the mean *r*-value for all the effect sizes was 0.45. The Fisher *rz* value transformed back into an *r*-value was 0.67 when all unknown zero values were excluded from analysis.

The combined *P*-value for the fixed effects method, showed *P* to be <0.0000001 [Rosenthal, 1991; Rosenthal and Rosnow, 1991]. The corresponding *Z*-value was equal to 8.70. The combined *P*-value using a random effects method calculated *P* to be <0.0000001. The corresponding *t*-value was equal to 9.39.

The file drawer *N*-value was 1,726. This represents the number of studies (that have not been retrieved) with a combined effect size of zero needed to nullify the results of the present analysis. This value of 1,726 is considered robust [Rosenthal, 1991]. Therefore, it is very unlikely that these significant results are nullified by the studies that were not retrieved for this analysis.

Independent Samples

Included in the meta-analysis were 11 studies that violated assumptions of independence. In particular these studies used the same animals for a number of different enrichment conditions. When these studies were excluded from analysis, the unweighted mean *r* for all independent effect sizes was 0.37. The weighted mean *r* was 0.34. The file drawer *N*-value was 1,022. This file drawer *N*-value is also considered robust [Rosenthal, 1991].

Measures of Spread

In Figure 1, a stem and leaf display shows the frequency distribution of the effect size *r* for all 63 experiments that examined the effect of enrichment on stereotypic behavior. Numbers in the right-hand column represent the hundredths place whereas the numbers in the left represent the tenths place. The data show mostly positive effects.

The homogeneity analysis suggested that the data had been drawn from different populations and was heterogeneous, with a $X^2 = 107.80$, $df = 63$, $P < 0.005$. A second homogeneity analysis was carried out on the known data (excluding assigned zero values). This homogeneity analysis showed the distribution to be homogeneous, or belonging to the same population with an $X^2 = 54.11$, $df = 46$, $P > 0.05$. These analyses suggests that all of the known effect sizes, regardless of enrichment type, species of animal, and form of stereotype, came from the same underlying population and that the heterogeneity was due to the unknown effect sizes of zero.

Contrasts

The planned contrasts analyzed the possible moderating variables of enrichment type, mammalian group, and form of stereotype performed by the captive animals. Each of these contrasts was carried out on all 63 effect-size statistics. Of these contrasts none was statistically significant. If the contrast size Z was ± 1.00, a second contrast was carried out excluding the effect size indices that were assigned a zero value. These analyses produced one statistically significant difference. In the variable of enrichment type, additional furniture did not reduce stereotypic behavior as much as the other forms of enrichment. Also, the contrast that examined the effect of food puzzles on frequency of stereotypic behavior was close to significant. This contrast suggested that food puzzles reduced stereotypic behavior more effectively than other types of enrichment (Table 2).

DISCUSSION

The current review systematically examined the influence different forms of environmental enrichment had on captive animals' negative stereotypic behaviors. The data overwhelmingly support the hypothesis that enrichment substantially reduces the frequency of stereotypic behavior exhibited by mammals living in zoo environments. The magnitude of the various mean effect size indices calculated for this analysis signifies a very large and significant treatment effect. The robustness of the results are most clearly understood when looking at the combined *P*-values for both fixed effects and random effects methods. The calculated P was < 0.0000001 when using both methods of calculation. In addition, the possibility of finding enough additional null studies to reach the file drawer *N*-value, which was calculated at 1,725, seems to be quite remote. The present analysis conclusively demonstrates

Stem	Leaf
.9	0, 2, 2, 3, 7, 8, 8
.8	0, 0, 1, 3, 3, 4, 4, 6, 9
.7	1, 1, 1, 1, 1, 3, 4, 6
.6	4, 4, 6, 9
.5	1, 1, 2, 3, 7, 8, 8, 9, 9
.4	5, 5, 8
.3	3, 5, 5
.2	
.1	5, 7
+0	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
?0	0, 0, 0
-0	0, 0, 4
-.1	
-.2	
-.3	
-.4	
-.5	
-.6	
-.7	
-.8	2

Fig. 1. Stem-and-leaf plot of the effect *r* of enrichment on stereotypic behavior in zoo animals.

that environmental enrichment can reduce stereotypic behavior in captive zoo mammals.

Enrichment techniques are typically evaluated as successful if the animals reduce their frequency or occurrence of stereotypic behavior while the enrichment is provided. This analysis found that although most forms of enrichment were

TABLE 2. Moderating variables

Variable type	Studies (n)	Contrast size (Z) All effect sizes	Contrast size (Z) Known effect sizes	Weighted mean rz All effect sizes	Weighted mean rz Known effect sizes
Form of enrichment					
Additional furniture	8	−1.58	−2.42*	0.398	0.507
Food puzzle	39	1.92	1.86	0.540	0.889
Scent	5	−1.25	−1.22	0.412	0.595
Taxonomic group					
Carnivores	37	1.28	.920	0.473	0.869
Marine mammals	10	−0.66		0.800	0.894
Primates	13	−0.261		0.426	0.809
Form of stereotype					
Pacing/pattern swimming	41	0.66		0.592	0.936
Abnormal behavior	10	−1.39	−0.65	0.335	0.698

Positive values suggest higher success in reducing stereotypic behavior; negative values suggest lower success in reducing stereotypic behavior.
*Significant at $P<0.05$.

successful and did reduce stereotype, none completely eliminated stereotypic behavior. This finding supports the literature review of Swaisgood and Shepherdson [2006]. Their review also found that although enrichment reduced stereotype it not abolish the behavior.

Contrasts

Of the three moderating variables that were examined, one moderating variable (type of enrichment) had two contrasts that were close to significant. When additional furniture was contrasted against all the other forms of enrichment the Z-value was −1.58, suggesting that additional furniture is less effective in reducing stereotypes than other types of enrichment. A second contrast was carried out on additional furniture when the assigned zero values excluded. In this contrast the Z-value was statistically significant and supports this suggestion. The second contrast that almost reached a significant value occurred when food puzzles were contrasted against all other forms of enrichment. This contrast had a Z-value of 1.92. Although not quite significant this value does suggest that food puzzles are often more effective in reducing stereotypic behavior than other forms of enrichment. A second contrast was carried out without the assigned zero values; this contrast had a Z-value of 1.86. Again, this value was close to a significant z and supports the suggestion that food puzzles reduce stereotypic behavior more effectively than other forms of enrichment.

Vote Count

The power of the meta-analytic techniques used in this review allowed for the examination of individual effects found in each primary study. The analysis showed that 90% of the effect sizes went in the predicted direction, with the captive animals engaging in less stereotypic behavior during the enrichment condition than in the baseline condition. After conducting the sensitive indices of combined *P*, overall direction, and individual effect size, a more traditional review method of "vote counting" was carried out [Lipsey and Wilson, 2001]. The vote counting analysis showed that 27% of the effects were statistically significant. This 27% significance rate is higher than what would be expected by chance alone. However, compared to the very small combined *P*-value of 0.0000001 and the substantial effect size *r* of 0.46, the 27% significance rate may seem a bit low. This is primarily because significance is reliant on sample size, and in enrichment literature sample size is consistently small [Swaisgood and Shepherdson, 2005]. The small sample sizes in the original literature reduced the authors' power to attribute statistical significance.

In addition to the "vote count" a mean *r* for the statistically significant articles was calculated to be 0.67. This value was then compared to the mean *r* for the non-significant articles with known values (all effect sizes other than the assigned zero values.) The calculated *r* for the non-significant articles was shown to be 0.60. (When all unknown zero values were included the *r*-value was calculated to be 0.38.) Clearly, the mean effects for both groups, significant ($r = 0.67$), and non-significant ($r = 0.60$) are not substantially different; rather it seems that in this literature statistical significance is due, primarily, to sample size.

Swaisgood and Shepherdson [2005] found that 53% of the 25 publications they analyzed reported a significant reduction in stereotypic behavior. This much higher percentage of statistical significance is due in part to their inclusion criteria, which only selected articles from three journals: *Animal Welfare*, *Applied Animal Behavior Sciences*, and *Zoo Biology*. These journals may over-represent the statistically significant findings due to their strict publishing guidelines. The inclusion criteria for the current meta-analysis focused on the original studies' methodologies rather than the publishing journal. As a result, these inclusion criteria may have provided a more comprehensive and less biased representation of the literature. The many nonsignificant findings reported in the Swaisgood and Shepherdson [2005] article and the current meta-analysis have substantially large effect sizes that should not be overlooked. "Vote counting" methods disregard nonsignificant findings even if the reported effect is substantial. Therefore, "vote counting" methods can underestimate the size of the effect. The results reported in this review suggest that when appropriate meta-analytic procedures are used, the effects of enrichment on stereotypic behavior are both consistent and substantial. In fact, all of the various ways of summarizing and analyzing the data arrive at the same conclusion: enrichment reduces stereotype. However, when using the less sophisticated conventional review procedure of "vote-counting" only 27% of the studies were significant, a value that could potentially mislead readers [Lipsey and Wilson, 2001]. In addition, the 53% significance rate found in the Swaisgood and Shepherdson [2005] article is also misleading when you understand that 90% of the studies in the literature actually report effects in the correct direction.

Regression to the Mean

Regression toward the mean is a universal phenomenon that occurs when two variables are imperfectly correlated and measured at more than one time point (for example, analyzing behavior during two conditions, baseline and enrichment.) For example, height and weight are two variables that are positively correlated but because they are imperfectly correlated the tallest people are not always the heaviest and the heaviest are not always the tallest [Lee and Smith, 2002]. Regression artifacts are an enormous problem when evaluating the effectiveness of a manipulation using quasi-experimental techniques that look for changes during two different observation periods [Campbell and Kenny, 1999]. These quasi-experimental designs dominate the field of enrichment. Typical enrichment studies provide animals that are participating in extreme behaviors (stereotypes) with an intervention and then re-evaluate the animals' behavior during the enrichment manipulation. Regression to the mean is always a possible explanation for differences observed in stereotypic behavior between baseline conditions and enrichment conditions. This is because the animals that are engaged in stereotypic behavior are exhibiting an extreme behavior pattern that may peak during the baseline condition. Therefore, the resulting decrease in stereotypic behavior during enrichment conditions may be due to the confounding effects of regression rather than the manipulation effects. Because of this potential problem, most of the studies presented in this analysis took certain precautions that reduced the risk of misinterpreting regression to the mean as a treatment effect. For the most part researchers used all the animals in the enclosure at the particular zoo, not just the animals participating in extreme amounts of stereotypic behavior. The animals expressing large amounts of stereotypic behavior, based on regression artifacts alone, would be expected to decrease their amount of stereotypic behavior. Additionally, the animals that carried out low levels of stereotypic behavior in the baseline condition should have carried out higher levels of stereotypic behavior during the enrichment intervention if regression to the mean were the sole explanation for the effect. In the original publications most animals provided with enrichment decreased their amount of stereotypic behavior regardless of the frequency of stereotypic activity exhibited before the intervention. These results suggest that regression to the mean did not account for the individual effects found in the original publications or for the large mean effect size found in this analysis.

In a perfect research world, experimenters would have enough animals to randomly assign them to conditions of enrichment and no enrichment. But until this more scientific approach is feasible researchers interested in studying environmental enrichment using baseline-enrichment (pre-post) designs should follow a few guidelines to be confident that the enrichment interventions (rather than regression or other artifacts) are responsible for the observed changes in behavior. First, researchers should report the stability of the stereotypic behavior during the baseline condition. If the stereotypic behavior remains consistently high over a substantial baseline period then consumers of the research may have more faith in the validity of the change. If the amount of time the animal spent in stereotypic behavior varied substantially and even seemed to increase during the baseline period, then researchers should consider regression artifacts as possible factors in the observed changes in behavior. Second, researchers should use all the animals within the enclosure as their subjects, rather than just the animals participating in extreme amounts of stereotypic behavior. Non-random selection (selection of the most extreme cases) tends to increase the amount of regression to the mean [Campbell and Kenny, 1999]. Fortunately for the field of zoo enrichment, it is often just as easy to provide all the animals within an enclosure with the enrichment manipulation. By using all the animals in the enclosure, regardless of their frequency of stereotype, researchers can more closely approximate a random sample. Third, researchers should collect data after the enrichment is removed. If the stereotype returns to baseline levels during such a phase, researchers can be more confident that regression to the mean is not the sole explanation for the changes in observed behavior. Generally, researchers should be cognizant of the effects of regression to the mean and provide reasonable arguments for why the changes observed in their data are not just regression artifacts.

Importance of Behavioral Analysis

Quantitative data collection provides researchers with a relatively unbiased evaluation of the effectiveness of the enrichment manipulation. In this field, large numbers of articles have been published that do not analyze the change or lack thereof in the animal's behavior. Rather, the articles suggest that the enrichment worked because the animals manipulated the device or in general seemed "happier" to the researcher [King, 1993; Camilla, 1995; Mason, 1995; Dickie, 1998; Rehling, 2000]. This unsystematic approach may lead frequently to the correct conclusion, but often people's perception of animal behavior can be incorrect. For example, many people interpret dolphins' "smile" as a sign that the animal is happy and is expressing emotions similar to those emotions of a smiling human. In actuality the smile shape of a dolphin's mouth is a physical trait created by the shape of the animals' jaws rather than an emotional expression [Masson and McCarthy, 1996]. When trying to improve an animal's enclosure or captive life by use of enrichment, anthropomorphizing can be a good starting point for coming up with novel enrichment ideas, but when evaluating the effectiveness of the enrichment manipulations quantitative assessment is needed to ensure accurate evaluation. Not all enrichment improves captive animal welfare; in fact some forms of intervention, believed to be enriching, can have a negative impact on the animal as shown by an increased amount of time in stereotype [Carlstead, 1991]. Carlstead [1991] provided four Fennec foxes with sand substrate that was supposed to be enriching. Unfortunately, the animals did not find the manipulation enriching and significantly increased the amount of time they spent in negative stereotypic behavior during the sand condition. In this example, behavioral measures were collected that clearly showed the negative impact the "enrichment" manipulation had on the animals' behavior. If these behavioral measures were not taken the researchers may have made erroneous conclusions about the effectiveness of the sand enrichment. Quantitative research on novel forms of enrichment can provide new ideas for enrichment programs as well as provide evaluative information. Articles that fail to include quantitative data often provide great new enrichment ideas [Poole and Kastelein, 1990; Grams, 1999; Rehling, 2000], but these enrichment manipulations need to be replicated with behavioral analysis to truly understand the effect of the enrichment on the behavior of animals.

Application to Zoo Routine

This review suggests that nearly any form of enrichment can and will reduce stereotypes in captive animals. This reduction in negative behavior may be a function of the novelty of the enrichment rather than the intrinsic qualities of the enrichment [Nash and Chilton, 1986; Kuczaj et al., 2002]. Theoretically, this is not a problem for the results reported in the individual articles nor it is a problem in this meta-analysis because providing animals with novelty can be a part of the enriching experience.

Yet, in zoo practice, all methods of enrichment may not continue to stimulate the animals and reduce negative behaviors. The studies in this analysis often collected data on the animals' behavior immediately after the introduction of the enrichment manipulation. A few of the studies mentioned that the animals continued to use the enrichment after data collection [Markowitz et al., 1995], but in general this information was not reported. Zoo personnel not only need to provide captive animals with appropriate enrichment, but also need to continually monitor the animals' response to the enrichment and the animals' general behavior when the enrichment is provided over a period of time. Certain types of enrichment may continue to stimulate animals of a particular species longer than other enrichment manipulations. For instance, food puzzles containing the animals' daily diet may remain stimulating for most species of carnivore, whereas a new piece of furniture (tree branch lying on the ground) may lose its enriching effects quickly.

CONCLUSION

Providing zoo researchers with the knowledge of powerful meta-analytic tools could have a substantial impact on zoo research. Future meta-analytic reviews could answer questions relating to enclosure size, diet, and reproductive success. The results of this particular meta-analysis strongly support the hypothesis that environmental enrichment has a positive behavioral influence on captive zoo animals. The presence of stereotypic behavior has long been understood to be an indicator of impoverished welfare [Wemelsfelder, 1993], and therefore methods to reduce or eliminate stereotypes from occurring in zoo animals are of primary importance. The current meta-analytic review suggests that regardless of the type of enrichment, the species of animal, or even the form of the stereotypic behavior, environmental enrichment reduces stereotypes and thereby improves captive zoo animal welfare.

ACKNOWLEDGMENTS

I am very grateful to M. Block and J.A. Hall for their helpful comments on the manuscript.

REFERENCES

Akers J, Schildkraut D. 1985. Regurgitation/reingestion and coprophagy in captive gorillas. Zoo Biol 4:99–109.

Altman J. 1999. Effects of inedible, manipulable objects on captive bears. J Appl Anim Welf Sci 2:123–32.

Anderson R, Wood J. 2001. Enrichment for giant Pacific octopuses: happy as a clam? J Appl Anim Welf Sci 4:157–68.

Angele C, Barber N, Plowman A. 1999. Elephant enrichment at Paignton Zoo Environmental Park. Proc Fourth Int Conf Environ Enrich 4: 311–5.

Ames A. 1999. Enclosure modification and design for captive polar bears. Proc Fourth Int Conf Environ Enrich 4:183–93.

Bashaw M, Bloomsmith M, Marr M, Maple T. 2003. To hunt or not to hunt? A feeding enrichment experiment with captive large felids. Zoo Biol 22:189–98.

Bauck L. 1998. Psittacian diets and behavioral enrichment. Semin Avian Exotic Pet Med 7:135–40.

Baxter E, Plowman A. 2001. The effect of increasing dietary fibre on feeding, rumination and oral stereotypies in captive giraffes (*Giraffa camelopardalis*). Anim Welf 10:281–90.

Broom D. 1983. Stereotypies as animal welfare indicators. In: Schmidt D, ed. Indicators relevant to farm animal welfare. The Hague: Martinus Nijhoff. p 8–87.

Bowkett A, Marshall A, Plowman A. 1999. Developing and evaluating an enrichment programme in an aging monkey house. Proc Fourth Int Conf Environ Enrich 4:304–10.

Brown D, Gold K. 1995. Effects of straw bedding on non-social and abnormal behavior of captive lowland gorillas (*Gorilla gorilla gorilla*). Proc Second Int Conf Environ Enrich 2:24–38.

Burghardt G, Ward B, Rosscoe R. 1996. Problem of reptile play: Environmental enrichment and play behavior in a captive Nile soft-shelled turtle, *Trionyz triunguis*. Zoo Biol 15:223–38.

Camilla N. 1995. Animal training as behavioral enrichment. Proc Second Int Conf Environ Enrich 2:165–7.

Campbell D, Kenny D. 1999. A primer on regression artifacts. New York: The Guilford Press.

Cardiff I. 1996. Assessing environmental enrichment for juvenile Jamaican boas Epicrates subflavus Stejneger, 1901. Dodo, J Wildl Preserv Trusts 32:155–62.

Carlstead K. 1991. Husbandry of the Fennec fox (*Fennecus zerda*) environmental conditions influencing stereotypic behavior. Int Zoo Yearb 30: 202–7.

Carlstead K, Brown J, Seidensticker J. 1993. Behavioral and adrenocortical responses to environmental changes in leopard cats (*Felis bengalensis*). Zoo Biol 12:321–31.

Carlstead K, Seidensticker J. 1991. Seasonal variation in stereotypic pacing in an American black bear *Ursus americanus*. Behav Processes 25: 155–61.

Carlstead K, Seidensticker J, Baldwin R. 1991. Environmental enrichment for zoo bears. Zoo Biol 10:3–16.

Chamove A. 1989. Environmental enrichment: a review. Anim Technol 40:155–78.

Chamove A, Anderson J, Morgan-Jones S, Jones S. 1982. Deep wood chip litter: hygiene, feeding, and behavioral enhancement in eight primate species. Int J Study of Anim Probl 3:308–18.

Cheng J. 2001. The effects of feeding pattern change on Malayan Sun bears. Proc Fifth Int Conf Environ Enrich 5:365–8.

Cohen J, Moore D. 2001. Behavioral enrichment of polar bear, *Ursus maritimus*, at the central park wildlife center. Proc Fifth Int Conf Environ Enrich 5:291.

Dantzer R. 1986. Behavioural, physiological and functional aspects of stereotyped behavior: a review and re-interpretation. J Anim Sci 62: 1776–86.

Dickie L. 1998. Environmental enrichment for Old World primates with reference to the primate collection at Edinburgh Zoo. Int Zoo Yearb 36: 131–9.

Duncan IJ. 1970. Frustration in the fowl. In: Freeman BM, Gordon RF, eds. Aspects of poultry behavior. Edinburgh, UK: British Poultry Science Ltd. p 15–31.

Fischbacher M, Schmid H. 1999. Feeding enrichment and stereotypic behavior in spectacled bears. Zoo Biol 18:363–71.

Forthman D, Elder S, Bakeman R, Kurkowski T, Noble C, Winslow S. 1992. Effect of feeding enrichment on behavior of three species of captive bears. Zoo Biol 11:187–95.

Gatcliffe E, Hudson C, Wehnelt S. 2002. Oral stereotypies in captive giraffes. The Federation of Zoos: Federation Research Newsletter 3: 5–6.

Garner J, Mason G, Smith R. 2003. Stereotypic route-tracing in experimentally caged songbirds correlates with general behavioral disinhibition. Anim Behav 66:711–27.

Gould E, Bres M. 1986. Regurgitation and reingestion in captive gorillas: description and intervention. Zoo Biol 5:241–50.

Grams K. 1999. Exhibitory and enrichment of North American river otters (*lontra Canadensis*) at the Arizona Sonora Desert Museum. Proc Fourth Int Conf Environ Enrich 4:196–209.

Grindrod J, Cleaver J. 2001. Environmental enrichment reduces the performance of stereotypic circling behavior in captive common seals (*Phoca vitulina*). Anim Welf 10:53–63.

Hall J, Rosenthal R. 1995. Interpreting and evaluating meta-analysis. Eval Health Prof 18: 393–407.

Hall JA, Tickle-Degnen L, Rosenthal R, Mosteller F. 1994. Hypotheses and problems in research synthesis. In: Cooper H, Hedges LV, eds. The handbook of research synthesis. New York: Russell Sage.

Hare V. 1995. Behavior modification of captive Alaska brown bears (*Ursus arctos gyas*) using automated food dispensers. Proc Second Int Conf Environ Enrich 2:39–61.

Hartmann M. 1997. A behavior-specific feeding technique for European wildcats (*Felis s.* silverstris). Proc Third Int Conf Environ Enrich 3: 182–90.

Hinde R. 1970. Animal behavior—A synthesis of ethology and comparative psychology. 2nd ed. New York: McGraw-Hill.

Hunter S, Bay M, Martin M, Hatfield J. 2002. Behavioral effects of environmental enrichment on harbor seals (*Phoca vitulina concolor*) and gray seals (*Halichoerus grypus*). Zoo Biol 21: 375–87.

Jenny S, Schmid H. 2002. Effect of feeding boxes on the behavior of stereotyping Amur tigers (*Panthera tigris altaica*) in Zurich Zoo, Zurich, Switzerland. Zoo Biol 21:573–84.

Kastelein R, Wiepkema P. 1989. A digging trough as occupational therapy for Pacific walruses (*Odobenus rosmarus divergens*) in human care. Aquat Mammals 15:9–17.

Kastelein R, Wiepkema P. 1988. The significance of training for the behavior of Steller sea lions (*Eumetopias jubata*) in human care. Aquat Mammals 14:39–41.

Keiper R. 1969. Causal factors of stereotypies in caged birds. Anim Behav 17:114–9.

King C. 1993. Environmental enrichment: is it for the birds? Zoo Biol 12:509–12.

Koene P, Kok J, Dekker J. 2001. Feeding whole carcasses in felids: is it enrichment for the individual? Proc Fifth Int Conf Environ Enrich 5:363.

Kolter L, Zander R. 1995. Potential and limitations of environmental enrichment in managing behavioral problems of polar bears. Proc Second Int Conf Environ Enrich 2: 131–41.

Kuczaj S, Lacinak T, Trone M, Solangi M, Ramos J. 2002. Keeping environmental enrichment enriching. Int J Comp Psychol 15:127–37.

Lee M, Smith G. 2002. Regression to the mean and football wagers. J Behav Decision Making 15: 329–42.

Lipsey M, Wilson D. 2001. Practical meta-analysis. Thousand Oaks, CA: Sage.

Lyons J, Young R. 1997. Effects of feeding regime manipulation upon behavior of felids at Edinburgh Zoo. R Zool Soc Scotl Annu Rep 1996:61–5.

Loeffler D. 1982. Regurgitation and reingestion in captive lowland gorillas. Zoo J 6:1–46.

Markowitz H, Aday C, Gavazzi A. 1995. Effectiveness of acoustic "prey": environmental enrichment of a captive African leopard (*Panthera pardus*). Zoo Biol 14:371–9.

Markowitz H, LaForse S. 1987. Artificial prey as behavioral enrichment devices for felines. Appl Anim Behav Sci 18:31–43.

Mason A. 1995. Fowl play at Edinburgh zoo. Shape Enrich 4:1–3.

Mason G. 1991. Stereotypies: a critical review. Anim Behav 41:1015–37.

Mason G. 1993. Age and context affect the stereotypies of mink. Behavior 127:191–229.

Mason GJ, Latham N. 2004. Can't stop, won't stop: is stereotypy a reliable animal welfare indicator. Anim Welf 13:S57–S69.

Masson J, McCarthy S. 1996. When elephants weep the emotional lives of animals. New York: Delta Publishing.

McPhee M. 2002. Intact carcasses as enrichment for large felids: effects on on- and off-exhibit behaviors. Zoo Biol 21:37–47.

Mellen J, MacPhee MS. 2001. Philosophy of environmental enrichment: past, present and future. Zoo Biol 20:211–26.

Meyer-Holzapfel M. 1968. Abnormal behavior in zoo animals. In: Fox MW, ed. Abnormal behavior in animals. London: Saunders.

Morimura N, Ueno Y. 1999. Influences on the feeding behavior of three mammals in the Maruyama Zoo: bears, elephants, and chimpanzees. J Appl Anim Welf Sci 2:169–86.

Nash L, Chilton S. 1986. Space or novelty? Effects of altered cage size on *Galago* behavior. Am J Primatol 10:37–49.

Newberry RC. 1995 Environmental enrichment: increasing the biological relevance of captive environments. Appl Anim Behav Sci 44:229–43.

Paquette D, Prescott J. 1988. Use of novel objects to enhance environments of captive chimpanzees. Zoo Biol 7:15–23.

Plowman A, Knowles L. 2001. Overcoming habituation in an enrichment program for tigers. Proc Fifth Int Conf Environ Enrich 5:263–7.

Poole T, Kastelein R. 1990. The role of training in the welfare of zoo mammals. Journal of the Association of British Wild Animal Keepers 17: 108–15.

Rehling M. 2000. Octopus prey puzzles. Shape Enrich 9:1–5.

Renner M, Lussier J. 2002. Environmental enrichment for the captive spectacled bear (*Tremarctos ornatus*). Pharmacol Biochem Behav 73:279–83.

Rooney M, Sleeman J. 1998. Effects of selected behavioral enrichment devices on behavior of Western lowland gorillas (*Gorilla gorilla gorilla*). J Appl Anim Welf Sci 1:339–51.

Rosenthal R. 1991. Meta-analytic procedures for social research. Newbury Park, CA: Sage.

Rosenthal R. 1995. Writing meta-analytic reviews. Psychol Bull 118:183–92.

Rosenthal R, Rosnow R. 1991. Essentials of behavioral research: methods and data analysis. New York: McGraw-Hill, Inc.

Ross S. 2002. The effect of a simple feeding enrichment strategy on the behavior of two Asian small-clawed otters (*Aonyx cinerea*). Aquat Mammals 28:113–20.

Rowell F. 1961. Displacement grooming in the Chaffinch. Animal Behavior 9:38–63.

Ruempler U. 1992. The Cologne Zoo diet for lowland gorillas to eliminate regurgitation and reingestion. Int Zoo Yearb 31:225–9.

Rushen J. 1984. Stereotyped behavior, adjunctive drinking and the feeding periods of tethered sows. Anim Behav 32:1059–67.

Rushen J, Lawrence AB, Terlouw CEM. 1993. The motivational basis of stereotypies. In: Lawrence AB, Rushen J, eds. Stereotypic animal behavior: fundamentals and applications to welfare. Wallingford, UK: CAB International. p 41–64.

Schiess-Meier M. 1997. Snow leopards (*Pantheria uncia*) using a new programmable play device. Proc Third Int Conf Environ Enrich 3:191–204.

Shepherdson D. 1989. Stereotypic behavior: what is it and how can it be eliminated or prevented? Ratel 16:100–5.

Shepherdson D, Brownback T, James A. 1989. A meal worm dispenser for the slender-tailed meerkat *Suricata suricatta* at London Zoo. Int Zoo Yearb 28:268–71.

Shepherdson D, Carlstead K, Mellen J, Seidensticker J. 1993. The influence of food presentation on the behavior of small cats in confined environments. Zoo Biol 12:203–16.

Small L. 2001. We can't be their jungle but we must try. Proc Fifth Int Conf Environ Enrich 5: 231–9.

Stoinski T, Daniel E, Maple T. 2000. A preliminary study of the behavioral effects of feeding enrichment on African elephants. Zoo Biol 19:485–93.
Swaisgood R, Shepherdson D. 2005. Scientific approaches to enrichment and stereotypies in zoo animals: what's been done and where should we go next? Zoo Biol 24:499–518.
Swaisgood R, Shepherdson D. 2006. Environmental enrichment as a strategy for mitigating stereotypies in zoo animals: a literature review and meta-analysis. In: Mason GJ, Rushen J, eds. Stereotypic animal behavior: fundamentals and applications to welfare. 2nd ed. Wallingford, UK: CAB International.
Swaisgood R, White A, Zhout X, Zhangt H, Zhangt G, Weit R, Hare V, Tepper E, Lindburg D. 2001. A quantitative assessment of the efficacy of an environmental enrichment program for giant pandas. Anim Behav 61:447–57.
Tepper E, Hare V, Swaisgood R, Lindburg D, Quan Z, Ripsky D, Hawk K, Hilpman G. 1999. Evaluating enrichment strategies with giant pandas at the San Diego Zoo. Proc Fourth Int Conf Environ Enrich 4:226–39.
Veninga S, Lemon J. 2001. Whole carcass feeding as a source of behavioral enrichment for African wild dogs (*Lycaon pictus*) in captivity at Western Plains Zoo, Dubbo. Proc Fifth Int Conf Environ Enrich 5:333–46.
Vivian M, Foster M, Sumpter C. 2001. The effect of a combination of environmental enrichment elements on the psychological well-being of captive chimpanzees (*Pan troglodyes*). Proc Fifth Int Conf Environ Enrich 5:131–43.
Wechsler B. 1992. Stereotypies and attentiveness to novel stimuli: a test in polar bears. Appl Anim Behav Sci 33:381–8.
Wehnelt S, Hudson C. 2002. Environmental enrichment for wild felids at Chester Zoo with emphasis on a new feeding enrichment. The Federation of Zoos: Federation Research Newsletter 3:10–1.
Wemelsfelder F. 1993. The concept of animal boredom and its relationship to stereotyped behavior. In: Lawrence A, Rushen J, eds. Stereotypic animal behavior: fundamentals and applications to welfare. Wallingford, UK: CAB International. p 65–95.
Wiard J. 1992. Reduction of regurgitation and reingestion (R&R) in lowland gorillas at the Oklahoma City Zoo. Gorilla Gazette 6:6–7.
Williams N, Chapman J, Plowman A. 1999. Olfactory enrichment for big cats, *Panthera leo persica* and *Panthera tigris sumatrae*. Proc Fourth Int Conf Environ Enrich 4:300–3.
Wood W. 1998. Interactions among environmental enrichment, viewing crowds, and zoo chimpanzees (*Pan troglodytes*). Zoo Biol 17:211–30.
Wozniak M. 1999. The effect of enrichment devices on stereotypic behavior in okapi (*Okapia johnstoni*). Proc Fourth Int Conf Environ Enrich 4:134–7.
Yanofsky R, Markowitz H. 1978. Changes in general behavior of two mandrills (*Papio sphinx*) concomitant with behavioral testing in the zoo. Psychol Rec 28:369–73.