ethology international journal of behavioural biology

Ethology

EDITORIAL

Sample Size in the Study of Behaviour

Michael Taborsky[1]

University of Bern, Bern, Switzerland

Correspondence
Michael Taborsky, Behavioural Ecology, Institute of Ecology and Evolution, University of Bern, Wohlenstrasse 50a, CH-3032 Hinterkappelen, Switzerland.
E-mail: michael.taborsky@iee.unibe.ch

[1] With this last contribution to Ethology as an editor I take the opportunity to thank all my co-editors for their marvellous cooperation. Over the last 10 yr in editing this journal, I had the wonderful opportunity to collaborate with (in alphabetical order) Gabriel Beckers, Johan Bolhuis, Jane Brockmann, Lee Drickamer, Luis Ebensberger, Scott Forbes, Susan Foster, Jean-Guy Godin, Alex Kacelnik, Bart Kempenaers, Janne Kotiaho, John Lazarus, Ronald Noë, Klaus Reinhold, Katharina Riebel, Scott Sakaluk, Jutta Schneider, Sarah Shettleworth, Lotta Sundström, Barbara Taborsky and David Zeh. Their serious dedication has rendered Ethology top regarding publication speed, which is the mainstay of its popularity and success. I am also grateful for the pleasant support of the publisher, particularly by Suzanne Albrecht. My special thanks go to Barbara for wisely steering the journal as managing editor through all the perils and pitfalls of editorial processing over all these years – without her commitment, efficiency and wit we could never have reached our goals.

doi: 10.1111/j.1439-0310.2010.01751.x

Abstract

The choice of an appropriate sample size for a study is a notoriously neglected topic in behavioural research, even though it is of utmost importance and the rules of action are more than clear – or are they? They may be clear if a formal power analysis is concerned. However, with the educated guesswork usually applied in behavioural studies there are various trade-offs, and the degrees of freedom are extensive. An analysis of 119 original studies haphazardly chosen from five leading behavioural journals suggests that the selected sample size reflects an influence of constraints more often than a rational optimization process. As predicted, field work involves greater samples than studies conducted in captivity, and invertebrates are used in greater numbers than vertebrates when the approach is similar. However, it seems to be less important for determining the number of subjects if the study employs observational or experimental means. This is surprising because in contrast to mere observations, experiments allow to reduce random variation in the data, which is an essential precondition for economizing on sample size. By pointing to inconsistent patterns the intention of this article is to induce thought and discussion among behavioural researchers on this crucial issue, where apparently neither standard procedures are applied nor conventions have yet been established. This is an issue of concern for authors, referees and editors alike.

An important decision at the outset of each scientific study of behaviour is the choice of an appropriate number of subjects. But what indeed is 'appropriate'? What might seem like a one-banana problem is a question with a hundred answers – at least!

Sample size determines the probability to reject a null hypothesis of no difference between populations when they are indeed different. In other words, sample size is a major determinant of statistical power, which is a measure of the probability that a study will detect a real difference in the data. The complement of statistical power is β or the Type-II error rate, which is the theoretical rate of failing to reject a false null hypothesis (i.e., power = 1-β). A low sample size raises the chances to commit a Type-II error, which might be called a 'false-negative result'.

There are different inferential approaches in science. Significance testing of a null hypothesis is only one of them, and not necessarily the best one as has been argued (e.g. Johnson 1999; Anderson et al. 2000, 2001; Howard et al. 2000; Hobbs & Hilborn 2006). Alternatives include information theoretic model comparisons, effect size statistics and Bayesian statistics (Burnham & Anderson 2002; McCarthy & Masters 2005; Nakagawa & Cuthill 2007; Stephens et al. 2007). What is regarded as adequate or best is a matter of convention, however, and significance testing of null hypotheses is by far the most common (in fact, almost sole) approach in behavioural studies. Regardless of the approach, however, the sample size used in a study determines the power to identify relationships among (statistical) populations. This can be illustrated by consideration of the essential components of significance tests: population variance, the Type-I error probability (α), i.e., the probability to wrongly reject a null hypothesis if the compared populations are indeed not different, the Type-II error probability (β) that is the complement of statistical power as outlined above, sample size and the effect size. The latter is a measure of the degree of difference between populations or 'the degree to which the phenomenon is present in the population' (Cohen 1988, p. 9). It is worth remembering here that the standard use of p-values in significance tests produces an apparent dichotomy that may actually obscure truth. Significance tests do not simply test the presence or absence of an effect; they are conditional on the effect size, or degree of departure of the effect from the null hypothesis. There are different ways to estimate effect size (e.g. Rosnow & Rosenthal 2003), but Cohen's d (Cohen 1962, 1988) is the most widely used. In its simplest form, for a two-tailed significance test of a difference between two independent groups

$$d = (m_1 - m_2)/\sigma \tag{1}$$

where m_1 and m_2 denote the arithmetic means of two samples and σ represents the population standard deviation. As usually we do not know these mean and variance estimates before conducting the study (which is when we have to decide about sample size), often conventions need to take the place of information (Cohen 1988).

Now it seems straightforward how to find the adequate sample size: There are conventions regarding α, β and even effect size, so the desired sample size can be calculated if we can assess the underlying distribution and have chosen the statistical test we wish to use. Various computer programs are available on the internet to calculate power statistics (e.g. Dupont & Plummer 1998; Lewis 2006; Faul et al. 2007; see links in Appendix). Naturally, the potential and value of this approach depends on the quality and plausibility of the conventions. The consensus is obviously strongest with regard to α, the acceptable level of the Type-I error probability, which in life sciences is almost universally 0.05, or 5%. This means that it is generally accepted that on average, in 1 of 20 cases a significance test wrongly rejects a null hypothesis even though the compared populations do not differ. There is more disagreement about the acceptable level of the Type-II error probability, because the potential damage caused by not finding a difference between populations when it actually exists is weighted much lower than the risk of seemingly finding it when it is not there. Often, researchers strive for β equalling 0.2, which corresponds to a power of 0.8 as recommended by Cohen (1988; but see Jennions & Møller 2003 for a survey revealing on average much lower power objectives in behavioural studies). In other words, it will be accepted that in one of five cases an existing difference will not be detected. The most disputed parameter in this calculation is effect size. If we do not know means and variance of the studied populations we must *define* a minimum ratio of the difference of their mean tendency over the variance (see Eq. 1) that we wish to detect with the given statistical power. Cohen (1988) suggested that if no information is available on population variance, standardized effect size classes can be used for small, medium and large effects, corresponding to 0.2, 0.5 and 0.8, respectively (but see Lenth 2001; Baguley 2004). If 'small' effects are expected, sample size would need to be increased to afford their detection.

As the power of a statistical analysis and hence the desired sample size strongly depends on the minimum detectable effect size, the prudent choice of this parameter is of crucial importance. If we are not happy with merely applying an arbitrary rule such as described above, we can either collect pilot data from the population(s) of interest, or rely on information from published evidence. For traits with frequency distributions measurable in ratio or interval scales, for example, it may be easy to make educated guesses about population mean and variance, and hence the calculation of adequate sample sizes will loose some of its subjectivity. Take a study of treatment effects on body size, for example, then the size distribution and variance of the underlying population may be well known and the only challenge to the researcher is to estimate the magnitude of the potential treatment effect on mean size. Choosing effect size for sample size optimization is usually less straightforward if behavioural measures are concerned, where the statistical parameters of the underlying population(s) are often unknown. This is probably one reason why statistical power analysis is apparently rarely used to determine sample size in behavioural research.

It should be stressed that even if prospective power analysis is a useful tool to determine sample size before data collection (Kraemer & Thiemann 1987; Cohen 1988; Steidl & Thomas 2001), retrospective power calculations using collected data will not provide meaningful information for the interpretation of results. This is because the 'observed power' is a 1:1 function of the p value, and the *post hoc* calculation of 'detectable effect size' leads to paradoxical conclusions (Hoenig & Heisey 2001). For a while, some behavioural journals requested post-experiment power calculations for non-significant test results, but fortunately this practice has ceased (Hoenig & Heisey 2001; Colegrave & Ruxton 2003; Nakagawa & Foster 2004).

There are other important considerations than statistical power when choosing sample size. Nothing comes without costs, so optimizing sample size involves trade-offs. Improving statistical power by opting for large samples will require more time and money to obtain the desired information. This aspect may be of minor importance in a study involving standard laboratory models such as fruit flies or certain rodent species bred for experimental purposes, or in studies of animals produced and kept in great numbers for commercial use. However, this may be of important concern if the access to animals is limited or difficult such as in many wildlife studies, or if the observational or experimental approach involves expensive provisions or time-consuming procedures. In behavioural studies, this question usually also involves an ethical dimension, because study subjects will be involved that may need to be disturbed, housed, manipulated or influenced in various ways to provide the required information. This is the reason that some behavioural journals such as 'Ethology' and 'Animal Behaviour' demand from their authors that sample sizes should be prudently chosen to avoid unnecessary suffering. There may be also serious methodological constraints, for example if the entire known, or captive, or available population is limited such as in endangered, vanishing or newly detected species. This should not stop us from performing research that may be particularly important *because* of such constraints, but rather we should be content with a small or moderate sample size at the expense of a higher β.

A simple example may help to illustrate the problem. Let us assume you wish to compare two independent samples, e.g. males and females of a species regarding their speed of locomotion during flight. You may have reasons to expect the underlying distribution of locomotion performance to be normal, you plan to use a Student's t-test for independent samples for analysis and you set the parameters of the statistical model to the conventional levels: $\alpha = 0.05$; $\beta = 0.2$; and Cohen's effect size $d = 0.5$. By reference to the respective t-distribution (tabled value at $\alpha = 0.05$ is 1.98) you arrive at a required total sample size of 128 for a two-tailed test, that is 64 males and 64 females. This may be practicable in an observational study such as described given that the animal densities are sufficient, but it may be hard to achieve and often impossible if a more complex approach is used, for example if animals have to be trained individually to learn a task in a social setting and the experimental conditions are demanding. If we aim for a more complex statistical model testing for several hypotheses at a time, our calculation may quickly reveal that several hundreds or thousands of individuals must be sampled to obtain the desired statistical power. Therefore it is no surprise that a meta-analysis of studies of evolution and ecology revealed that the average sample sizes needed to conclude that a particular relationship is absent with a power of 80% and alpha = 0.05 (two-tailed) would be much larger than those usually chosen (Møller & Jennions 2002).

There is another important aspect that must be considered when planning a study: we should determine which population exactly is the target of inference. If we are interested in the vigilance of great tits at winter feeders in dependence of the interspecific composition of feeding flocks, for example, our target of inference may be the great tits of a geographical region, a population living under particular climatic conditions, great tits living in urban or rural environments or under any other specific ecological condition, males, females or juveniles, great tits at feeders in early, mid or late winter, great tits of a certain latitude or altitude, or 'the great tit' in general under the widest range of conditions. The target of inference determines the sampling approach and naturally also the projected sample size. The primary 'unit of study' is another important issue requiring choice that will influence sample size considerations. The unit of scrutiny is mostly, but not always, individuals; it may be, for example, dyads, groups, certain types of interactions, behaviours displayed in certain situations (e.g. in a problem solving context), time-series of treatments, etc.

Given the potential limitations in behavioural research to attain sample sizes matching the desired statistical power, is there a way out of the dilemma? Indeed, sample size is not the only determinant of statistical power. Power depends also on effect size, i.e. the 'real difference' between the populations we wish to compare, and the amount of random variation. We have discussed the intricacies of determining effect size above. A question of great practical utility regards the potential to influence the variance in the data we wish to collect. The smaller the variance, the lower the number of samples we need to obtain the desired statistical power. Therefore, alternatives to boosting sample size include, for instance, to improve the experimental design (McClelland 1997, 2000; Quinn & Keough 2002; Ruxton & Colegrave 2006), to reduce the variance in the data by prudent sampling and maximizing accuracy in parameter estimation (Schwagmeyer & Mock 1997; Maxwell et al. 2008; e.g. by using repeated measures instead of randomized groups; Still 1982), and to avoid omnibus tests by asking specific questions and selecting adequate tests (Baguley 2004; Maxwell 2004; Nakagawa 2004; Brauer & McClelland 2005; Ruxton & Beauchamp 2008). In the light of these possibilities and the constraints mentioned above, several factors should influence the choice of sample size in behavioural studies. Most importantly, in contrast to purely observational means an experimental approach should afford to keep the majority of confounding factors constant while permitting to systematically vary the independent factor(s) of interest. Therefore, we should predict that in general, experimental studies show smaller sample sizes than observational surveys. In addition, if animals are studied in the field, the number of potential confounding factors usually exceeds those affecting investigations in captive settings. Therefore, field studies should be predicted to use larger sample sizes than laboratory studies. Regarding the effect of constraints, it is more ethical, easier, cheaper and less time consuming to use large samples when, say, studying invertebrates in the laboratory than when conducting a similar study with primates. Therefore, larger sample sizes should be used in studies of animals that can be easily held, bred and manipulated, and where ethical concern is of lower priority than in other species.

Let us check these predictions with recent publications in ethology. Is the choice of sample size in behavioural research indeed triggered by (1) the approach used (experimental or observational), (2) the location of a study (field or captive site) and (3) taxon-specific economical and ethical constraints? To answer these questions, I analysed the sample sizes of a haphazard sample of articles published in 2009 in 'Ethology' (from the January, April, July and October issues), and from matching issues, respectively, of two other general behavioural journals ('Behaviour' and 'Animal Behaviour') and of two journals more specialized in the ultimate and proximate mechanisms of behaviour, respectively ('Behavioral Ecology' and 'Behavioural Processes'). Of these issues, all full-length original articles were considered that reported adequate data and sampling information ($N = 119$ studies that fulfilled these criteria, reported in 107 articles; Table 1).

Univariate ANOVA confirmed the predicted pattern (the dependent variable was log-transformed to comply with test requirements; medians and two-tailed error probabilities are given; analyses were performed with SPSS v. 17.0): observational studies use on average greater sample sizes than those employing an experimental approach (34 vs. 20.9; $p = 0.01$; $df = 1$; $N = 39$ observational and 80 experimental studies, respectively), and field studies involve larger samples than research performed in a captive setting (32 vs. 18; $p < 0.001$; $df = 1$; $N = 50$ field and 65 captive studies; four human studies utilizing questionnaire-type measures were not included). When both 'site' and 'approach' were included in the analysis, the factor 'site' was significant ($p = 0.007$; $df = 1$), but 'approach' was not ($p = 0.55$; $df = 1$) and neither was the interaction between these two factors ($p = 0.38$; $df = 1$). This suggests that the significant effect of 'approach' when tested separately was actually due to different sample sizes used in field and laboratory studies, because these conditions are not independent. The model including all three main factors approach, site and the taxonomic groups 'invertebrates' ($N = 33$; $\tilde{x} = 32.5$), 'poikilothermic vertebrates' ($N = 11$; $\tilde{x} = 15$), 'birds' ($N = 35$; $\tilde{x} = 24.7$), and 'mammals' ($N = 36$; $\tilde{x} = 21$) suggested that the taxonomic group affected sample size ($p = 0.005$; $df = 3$), with the greatest sample sizes used in studies of invertebrates, as expected (questionnaire-type studies were not included; non-significant interaction terms were removed stepwise).

These results reveal that by and large, ethologists are using criteria other than – or in addition to – statistical power to determine sample size. When

Table 1: Sample sizes used in a haphazard sample of all full-length original articles published in (a) the January, April, July and October issues of 'Ethology' in 2009, and (b) matching issues of four other behavioural journals, respectively: Behaviour (January), Animal Behaviour (April), Behavioral Ecology (July) and Behavioural Processes (October). Sample sizes refer to independent units (usually individuals) per treatment (combined by arithmetic operators) and/or test (separated by semicolons). If a multitude of tests with varying sample sizes were reported, the range is given (e.g. 4–18). The major statistical procedures used in the studies are also reported. As most studies involved more than one test and usually sample sizes differed among tests and treatments, a mean estimate per study was calculated (columns 'Calculation of N' and the resulting 'N for analysis'; the latter was used for the analysis reported in the text). Some articles reported two substantial studies [e.g. captive and field, or observational (obs.) and experimental (exp.)], which are listed in separate lines

Research issue	Genus	Taxonomic group	Approach	Site	Reported N	Statistics	Reference	Calculation of N	N for analysis
(a)									
January									
Parental care	*Nicrophorus*	Insect	exp.	captive	15 + 15; 15 + 16; 18	ANOVA	Suzuki & Nagano (2009)	(15 + 15.5 + 18)/3	16.2
Running performance	*Sceloporus*	Reptile	obs.	field	166	–	Cooper et al. (2009a)	166	166
			exp.	captive	13 + 11; 6 + 6; 3 + 3	ANOVA		(12 + 6 + 3)/3	7
Parasite–host competition	*Cuculus*	Bird	exp.	field	10 + 9	GLMM	Grim et al. (2009)	(10 + 9)/2	9.5
Species recognition	*Taenopygia*	Bird	exp.	captive	12	GLMM	Campbell et al. (2009)	12	12
Foraging interference	*Esox*	Fish	exp.	captive	8 × 6	ANOVA	Nilsson et al. (2009)	8	8
Behavioural devel.	*Pundamilia*	Fish	exp.	captive	4–18	GLMM	Verzijden et al. (2009)	(4 + 18)/2	11
Parental care	*Ficedula*	Bird	exp.	field	24 + 24; 29 + 19	ANOVA	Wiebe & Slagsvold (2009)	(24 + 24 + 29 + 19)/4	24
Social organization	*Cynomis*	Mammal	obs.	field	20	Fisher's comb. prob. test; GLM	Verdolin & Slobodchikoff (2009)	20	20
Grooming reciprocation	*Cebus*	Primate	obs.	captive	18 (7)	Survival anal.; mixed model regr.	Schino et al. (2009)	18	18
Duetting	*Pipilo*	Bird	obs.	field	90; 57; 60 (8; 9)	PCA; Discrim. Funct. Anal.	Benedict & McEntee (2009)	(90 + 57 + 60)/3	69
Partner preference	*Mus*	Mammal	exp.	captive	100	GLM	Busquet et al. (2009)	100	100
Sperm expenditure	*Notophthalmus*	Amphibian	exp.	captive	10	ANOVA	Takahashi & Parris (2009)	10	10
April									
Aesthetic preference	*Homo*	Human	obs.	quest.	83 + 83	PCA; Discrim. Funct. Anal.	Maresova et al. (2009)	83 + 83	166
Division of labour	*Lasioglossum*	Insect	exp.	captive	39 + 37	MANOVA	Holbrook et al. (2009)	(39 + 37)/2	38
Risk assessment	*Dendrobates*	Amphibian	exp.	field	37; 10; 14	ANOVA; t-test; non-param. tests	Cooper et al. (2009b)	(37 + 10 + 14)/3	20.3
Collective movement	*Equus*	Mammal	obs.	captive	12 + 6	Non-param. tests	Bourjade et al. (2009)	12 + 6	18
Aggressive behaviour	*Pardosa*	Spider	exp.	captive	21 + 24 + 20 + 19 + 19	Kruskal–Wallis ANOVA	Hoefler et al. (2009)	(21 + 24 + 20 + 19 + 19)/5	20.6
Sex recognition	*Syngnathus*	Fish	exp.	captive	17 + 12; 16–27	Wilcoxon signed-ranks test	Ratterman et al. (2009)	((17 + 12)/2 + (16 + 27)/2)/2	18
Sequential resource use	*Coenobita*	Crustacean	exp.	captive	6; 5; 13	t-test; binomial test	Lewis & Rotjan (2009)	(6 + 5 + 13)/3	8
			exp.	field	348 + 265	Contingency table test; ANOVA		(348 + 265)/2	306.5
Handling stress	*Parus*	Bird	exp.	field	236	GLMM	Fucikova et al. (2009)	236	236

Table 1: (Continued).

Research issue	Genus	Taxonomic group	Approach	Site	Reported N	Statistics	Reference	Calculation of N	N for analysis
			exp.	captive	37	GLMM		37	37
Mate attraction	*Lycosa*	Spider	exp.	captive	20	Non-param. tests	Fernandez-Montraveta & Cuadrado (2009)	20	20
			exp.	field	7 + 6 + 7; 6–16	Non-param. tests		((7 + 6 + 7)/3 + (6 + 16)/2)/2	8.8
Social organization	*Lontra*	Mammal	obs.	captive	15	t-test; ANOVA; Mantel test	Hansen et al. (2009)	15	15
July									
Migration	*Otis*	Bird	obs.	field	65 + 68	Non-param. tests	Palacin et al. (2009)	(65 + 68)/2	66.5
Symmetry and generosity	*Homo*	Human	exp.	quest.	106 + 82	t-test; non-param. tests	Zaatari et al. (2009)	106 + 82	188
Sexual cannibalism	*Argiope*	Spider	exp.	captive	20–43	ANOVA; G-test	Nessler et al. (2009)	(20 + 43)/2	31.5
Winter foraging	*Peromyscus*	Mammal	exp.	field	10 × 5	ANOVA	Orrock & Danielson (2009)	10	10
Reconciliation	*Macaca*	Primate	obs.	captive	8	GLMM	Patzelt et al. (2009)	8	8
Competitive ability	*Macaca*	Primate	exp.	captive	187	Pearson correlation	Bissonnette et al. (2009)	187	187
Offspring sex ratio	*Poecilia*	Fish	exp.	captive	25; 37	t-test; G-test; logistic regression	Karino & Sato (2009)	(25 + 37)/2	31
Spatial competition	*Argiope*	Spider	exp.	captive	15 × 5	ANOVA	Rao (2009)	15	15
Mating flight costs	*Apis*	Insect	exp.	captive	20 + 21 + 39; 6 × 3	ANOVA; ANCOVA	Hayworth et al. (2009)	(20 + 21 + 39 + 6)/4	21.5
Ideal free distribution	*Columba*	Bird	obs.	field	4 × 7	REML	Morand-Ferron et al. (2009)	28	28
October									
Contextual vocalization	*Canis*	Mammal	obs.	captive	32	t-test	Taylor et al. (2009)	32	32
	Homo	Human	exp.	quest.	33	ANOVA		33	33
Microhabitat tenacity	*Latrodectus*	Spider	exp.	field	60; 44	GLM; Fisher's exact test; surv. anal.	Salomon (2009)	(60 + 44)/2	52
			exp.	captive	48	GLM		48	48
Alarm call response	*Ochotona*	Mammal	exp.	field	30; 30; 30	G-test	Trefry & Hik (2009)	30	30
Mate choice	*Hirundo*	Bird	obs.	field	100; 71	GLMM	Kojima et al. (2009)	(100 + 71)/2	85.5
Collective movement	*Bos*	Mammal	obs.	captive	26; 46	GLM	Ramseyer et al. (2009)	(26 + 46)/2	36
Signal inheritance	*Poecilia*	Fish	exp.	captive	458	ANOVA; ANCOVA; regression	Loveless et al. (2009)	458	458
Social organization	*Macroscelides*	Mammal	obs.	field	8–33	LMM; ANOVA	Schubert et al. (2009)	(8 + 33)/2	20.5
Anxiety and reconciliation	*Macaca*	Primate	obs.	field	41	LMM; GLM; GLMM	Majolo et al. (2009a)	41	41
Dominance effects	*Gallus*	Bird	exp.	captive	15	Friedman ANOVA; Wilcoxon test	Wilson et al. (2009)	15	15

Table 1: (Continued).

Research issue	Genus	Taxonomic group	Approach	Site	Reported N	Statistics	Reference	Calculation of N	N for analysis
(b)									
January: Behaviour									
Sperm competition	*Austropotamobius*	Crustacean	exp.	captive	72; 73; 74; 111; 49; 42; 18 + 15; 25 + 33; 55 + 56	bin.logist.regr., GLM, Non-par.tests	Galeotti et al. (2009)	(72 + 73 + 74 + 111 + 49 + 42 +((18 + 15)/2)+((25 + 33)/2)+((55 + 56)/2))/9	58
Social organization	*Setonix*	Mammal	obs.	field	21 + 40; 21; 17; 20 + 19; 20; 28; 15 + 21; 15 + 20	Pearson corr.; non-param. tests	McLean et al. (2009)	(((21 + 40)/2)+21 + 17 +((20 + 19)/2)+20 + 28 + ((15 + 21)/2)+((15 + 20)/2))/8	21.4
Group size effects	*Macaca*	Primate	obs.	field	8 + 20	ANCOVA; GLMM	Majolo et al. (2009b)	(8 + 20)/2	14
Bruce effect	*Otomys*	Mammal	exp.	captive	10 × 5; 10 × 5	GLM; regression; Fisher's exact test	Pillay & Kinahan (2009)	10	10
April: Animal Behaviour									
Innate releasing mechanism	*Larus*	Bird	exp.	field	25; 24	GLMM; LMM	ten Cate et al. (2009)	(25 + 24)/2	24.5
Behavioural syndromes	*Ficedula*	Bird	exp.	field	16–26	Non-parametric tests	Garamszegi et al. (2009)	(16 + 26)/2	21
Antipredator response	*Schizocosa*	Spider	exp.	captive	10; 16–21; 23 + 20	ANOVA	Lohrey et al. (2009)	(10 + ((16 + 21)/2)+ ((23 + 20)/2))/3	16.7
Dominance and mating	*Phoxinus*	Fish	exp.	captive	15	LMM; regression	Jacob et al. (2009)	15	15
Dispersal and philopatry	*Gorilla*	Primate	obs.	field	44; 116	Fisher's exact test; GLMM	Robbins et al. (2009)	(44 + 116)/2	80
Social organization	*Macropus*	Mammal	obs.	field	38	HWIs; simulation model	Carter et al. (2009)	38	38
Antipredator response	*Sepia*	Invertebrate	exp.	captive	20 + 13	Fisher's exact test	Langridge (2009)	(20 + 13)/2	16.5
Food preference	*Sturnus*	Bird	exp.	captive	9 + 9; 20	ANOVA	Freidin et al. (2009)	((9 + 9)/2 + 20)/2	14.5
Agression	*Sialia*	Bird	obs.	field	26; 111	GLMM; GLM	Dickinson et al. (2009)	(26 + 111)/2	68.5
Dominance and paternity	*Pan*	Primate	obs.	field	34	ANOVA; regression; non-param. tests	Wroblewski et al. (2009)	34	34
Ejaculate effects	*Gryllus*	Insect	exp.	captive	22 + 23; 41 + 48; 27 + 34	Non-parametric tests; t-test	Green & Tregenza (2009)	((22 + 23)/2 + (41 + 48)/2 + (27 + 34)/2)/3	32.5
Food shortage response	*Gallus*	Bird	exp.	captive	26 + 27; 18 + 18	GLM	Lindqvist et al. (2009)	((26 + 27)/2 + (18 + 18)/2)/2	22.3
Attraction to volatiles	*Aphidius*	Insect	exp.	captive	146; 124	PCA; GLM	Pareja et al. (2009)	(146 + 124)/2	135
Temporal polyethism	*Pheidole*	Insect	exp.	captive	12 + 13; 4 + 4	ANOVA; t-test	Muscedere et al. (2009)	((12 + 13)/2 + (4 + 4)/2)/2	8.3

Table 1: (Continued).

Research issue	Genus	Taxonomic group	Approach	Site	Reported N	Statistics	Reference	Calculation of N	N for analysis
Social organization	*Ovis*	Mammal	obs.	field	8; 46 + 53; 25 + 49; 25 + 25	LMM; non-parametric tests	Meldrum & Ruckstuhl (2009)	(8 + (46 + 53)/2 + (25 +49)/2 + (25 + 25)/2)/4	29.9
Sperm competition	*Cadra*	Insect	exp.	captive	37 + 31; 33 + 33	Kruskal-Wallis ANOVA; GLM	McNamara et al. (2009)	((37 + 31)/2 + (33 + 33)/2)/2	33.5
Multimodal signalling	*Pardosa*	Spider	exp.	captive	20–35	ANOVA	Rypstra et al. (2009)	(20 + 35)/2	27.5
Social organization	*Panthera*	Mammal	obs.	field	140	Regression; GLM; chi-square test	VanderWaal et al. (2009)	140	140
Heart rate response	*Anser*	Bird	obs.	field	24	GLMM; REML	Wascher et al. (2009)	24	24
Social organization	*Orcinus*	Mammal	obs.	field	37 + 31 + 89	Bayesian mixture model	Parsons et al. (2009)	(37 + 31 + 89)/3	52.3
Vocal performance	*Melospiza*	Bird	obs.	field	31; 21; 22	GLM	Ballentine (2009)	(31 + 21 + 22)/3	24.7
July: Behavioral Ecology									
Diet selection	*Ateles*	Primate	obs.	field	15	Regression	Felton et al. (2009)	15	15
Brood parasitism	*Molothrus*	Bird	exp.	field	39–41; 29–30; 22–23; 13–14; 16; 3;	Fisher's exact test	Strausberger & Rothstein (2009)	((39 + 41)/2 + (29 + 30)/2 + (22 + 23)/2 + (13 + 14)/2 + 16 + 3)/6	20.8
			obs.	field	7–11 + 7 + 6; 113 + 74 + 23 + 86 + 114 + 178 + 50 + 529	Fisher's exact test		(((7 + 11)/2 + 7 + 6)/3 + (113 + 74 + 23 + 86 + 114 + 178 + 50 + 529)/8)/2	76.6
Diet, courtship and mating	*Pardosa*	Spider	exp.	captive	137 + 135; 125 + 100; 22; 44 + 32; 18 + 12; 32	ANOVA; ANCOVA; PCA	Lomborg & Toft (2009)	((137 + 135)/2 + (125 + 100)/2 + 22 + (44 + 32)/2 + (18 + 12)/2 + 32)/6	59.3
Vigilance and predation	*Perisoreus*	Bird	obs.	field	31 + 35; 20 + 13	GLM; LMM	Griesser & Nystrand (2009)	((31 + 35)/2 + (20 + 13)/2)/2	24.8
			exp.	field	12 × 3	LMM		12	12
Aesthetic preference	*Homo*	Human	exp.	quest.	100	Multiple regression	Donohoe et al. (2009)	100	100
Phonotaxis	*Teleogryllus*	Insect	exp.	captive	57 + 61 + 55 + 57; 24 + 28 + 55 + 57	ANOVA	Tinghitella et al. (2009)	((57 + 61 + 55 + 57)/4 +(24 + 28 +55 + 57)/4)/2	49.3
Offspring competition	*Mungos*	Mammal	obs.	field	82; 59; 180; 131; 39	LMM; GLMM	Hodge et al. (2009)	(82 + 59 + 180 + 131 + 39)/5	98.2

Table 1: (Continued).

Research issue	Genus	Taxonomic group	Approach	Site	Reported N	Statistics	Reference	Calculation of N	N for analysis
Parental care	*Morus*	Bird	exp.	field	26 + 49	Multilevel modeling; GLM	Bijleveld & Mullers (2009)	(26 + 49)/2	37.5
Alarm call response	3 passerines	Bird	exp.	field	12 × 9	Cochran Q test	Magrath et al. (2009)	12	12
Alarm calling			obs.	field	11 × 3	Cochran Q test; non-param. tests		11	11
Competitive fertilization	*Drosophila*	Insect	exp.	captive	71 + 83; 55	GLM; ANCOVA	Polak & Simmons (2009)	((71 + 83)/2 + 55)/2	66
Brood size	*Copidosoma*	Insect	exp.	captive	65 + 51 + 13; 18 + 30; 40; 34	ANOVA; GLM	Segoli et al. (2009)	((65 + 51 + 13)/3 + (18 + 30)/2 + 40 + 34)/4	35.3
			obs.	field	101	Regression		101	101
Egg allocation	*Syngnathus*	Fish	exp.	captive	5 × 6 × 2;	ANOVA	Silva et al. (2009)	5	5
Dominance hierarchy	*Polistes*	Insect	exp.	field	53	Rank correlation; GLM	Zanette & Field (2009)	53	53
Behavioural syndrome	*Gromphadorhina*	Insect	exp.	captive	17; 70; 39	Pearson correlation	Logue et al. (2009)	(17 + 70 + 39)/3	42
Nest defence	*Strix*	Bird	obs.	field	244; 127; 17	LMM; path analysis; GLM	Kontiainen et al. (2009)	(244 + 127 + 17)/3	129.3
Plumage dimorphism	*Limosa*	Bird	obs.	field	74 + 72; 57 + 60; 61 + 79	GLM; LMM	Schroeder et al. (2009)	((74 + 72)/2 + (57 + 60)/2 + (61 + 79)/2)/3	67.2
Signal design	*Nephila*	Spider	exp.	field	23 + 18 + 19 + 19 + 18	ANOVA; regression; non-param. tests	Fan et al. (2009)	(23 + 18 + 19 + 19 + 18)/5	19.4
Colony switch	*Apis*	Insect	exp.	captive	4 × 3	Chi-square test	Chapman et al. (2009)	12	12
Begging and predation	*Manorina*	Bird	exp.	field	21 × 2 × 4	PCA; survival analysis	McDonald et al. (2009)	21 × 4	84
Behavioural syndrome	*Sturnus*	Bird	obs.	field	320 + 304; 42; 24	PCA; LMM	Minderman et al. (2009)	((320 + 304)/2 + 42 + 24)/3	126
Alloparental care	*Physeter*	Mammal	obs.	field	23	t-test	Gero et al. (2009)	23	23
Mate choice	*Pirata*	Spider	exp.	captive	135 + 40 + 94 + 49; 28 + 25 + 75 + 70; 19 + 16 + 27 + 20 + 17 + 11	ANOVA; exact tests; GLMM	Eraly et al. (2009)	((135 + 40 + 94 + 49)/4 + (28 + 25 + 75 + 70)/4 + (19 + 16 + 27 + 20 + 17 + 11)/6)/3	49.1
Secondary seed dispersal	*Falco*	Bird	exp.	captive	5 × 6	Wilcoxon signed-ranks test; ANOVA	Padilla & Nogales (2009)	5 × 6	30
Parental care	*Pygoscelis*	Bird	obs.	field	14 + 18; 18 + 18	GLM; GLMM	Beaulieu et al. (2009)	((14 + 18)/2 + (18 + 18)/2)/2	17
Diet selection	*Macrobiotus*	Invertebrate	exp.	captive	88; 15; 3; 16	Regression; t-test; ANOVA	Hohberg & Traunspurger (2009)	(88 + 15 + 3 + 16)/4	30.5
Mating tactic	*Rabidosa*	Spider	exp.	captive	92; 40 + 42	ANOVA; regression; non-param. tests	Wilgers et al. (2009)	(92 + (40 + 42)/2)/2	66.5

Table 1: (Continued).

Research issue	Genus	Taxonomic group	Approach	Site	Reported N	Statistics	Reference	Calculation of N	N for analysis
Song performance	*Junco*	Bird	obs.	field	97	ANOVA; Pearson correlation	Cardoso et al. (2009)	97	97
October: Behavioural Processes									
Response to deception	*Canis*	Mammal	exp.	captive	7	t-test; ANOVA	Petter et al. (2009)	7	7
Mate choice	*Mus*	Mammal	exp.	captive	48	chi-square test; t-test; ANOVA	Costello et al. (2009)	48	48
Conditioned reinforcement	*Columba*	Bird	exp.	captive	4 × 4	ANOVA; t-test	Gipson et al. (2009)	4 × 4	16
Vocal matching	*Aratinga*	Bird	exp.	field	169; 71;	GLM; LMM	Balsby & Bradbury (2009)	(169 + 71)/2	120
			exp.	captive	14	LMM		14	14
Picture processing	*Cebus*	Primate	exp.	captive	4	ANOVA	Truppa et al. (2009)	4	4
Freezing behaviour	2 genera	Spider	exp.	captive	17 × 3; 22 × 3; 13; 16 × 3	ANOVA; non-parametric tests	Chelini et al. (2009)	(17 + 22 + 13 + 16)/4	17
Stimulus matching	*Pan*	Primate	exp.	captive	1	Binomial test; Fisher's exact test	Martinez & Matsuzawa (2009)	1	1
	Homo	Human	exp.	'captive'	10	t-test; binomial test; ANOVA		10	10
Probability matching	*Homo*	Human	exp.	'captive'	6 × 10 × 100; 6 × 10 × 100	t-test	Kangas et al. (2009)	(6 + 6)/2	6
Social preference	*Passer*	Bird	obs.	captive	21 + 23 + 17	Mantel test	Tóth et al. (2009)	(21 + 23 + 17)/3	20.3
Reinforcement tracking	*Canis*	Mammal	exp.	captive	8	t-test	Gilbert-Norton et al. (2009)	8	8
Diet selection	*Ovis*	Mammal	exp.	captive	8 × 4	LMM; regression	Lisonbee et al. (2009)	8	8
Cue-dependent navigation	*Rattus*	Mammal	exp.	captive	15; 8 + 7	ANOVA	Harvey et al. (2009)	(15 + (8 + 7)/2)/2	11.2
Call discrimination	*Canis*	Mammal	exp.	captive	30 × 3; 20	Non-parametric tests	Molnár et al.	((30 × 3)/3 + 20)/2	25
Reactivity to handling	*Equus*	Mammal	exp.	captive	6 × 2	t-test; GLM	Fureix et al. (2009)	6	6

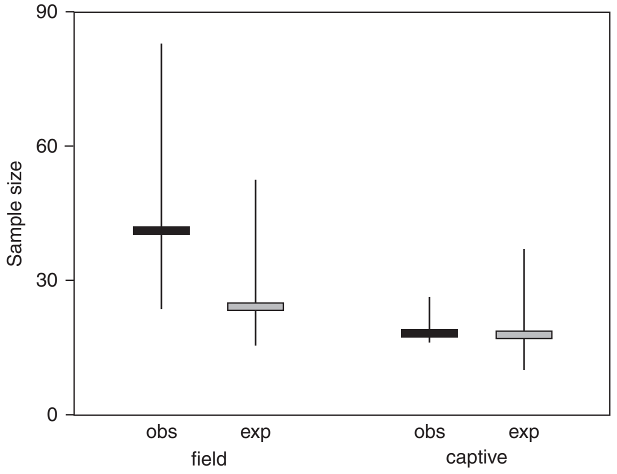

Fig. 1: Medians and interquartile ranges of sample sizes used in a total of 119 studies included in Table 1. Studies were performed in the field or in a captive setting, and with an observational (obs.) or experimental (exp.) approach.

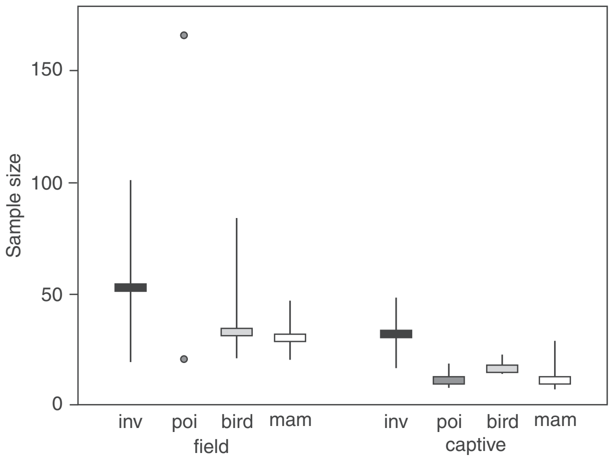

Fig. 2: Medians and interquartile ranges of sample sizes used in studies of invertebrates (inv), poikilothermic vertebrates (poi), birds (bird) and mammals (mam). The original values are given in Table 1. There were only two field studies of poikilothermic vertebrates in the sample, which are represented by filled circles.

studying animals under captive conditions the number of subjects is apparently reduced on average to approx. 60% of the numbers used in field studies. However, the difference between observational and experimental approaches is less clear (Fig. 1); this cannot be resolved conclusively because of the non-independence of the factors 'site' and 'approach' in the data set, but the analysis showed a significant effect of 'site' and not of 'approach' when both were included. This is surprising because if anything, an experimental approach allows to control confounding factors and therefore experiments are the method of choice if the intention is to reduce variance in the data to economize on sample size. Also, the range of sample sizes used is enormous (e.g. 1–458 in the sample of experimental studies performed in captivity). Ethical concerns and other constraints are presumably responsible for the fact that sample sizes in studies involving vertebrates are lower than when invertebrates are investigated (Fig. 2; $\tilde{x} = 21.8$ for all vertebrates combined excluding questionnaire-type studies, vs. $\tilde{x} = 32.5$ for all invertebrates), even though with a divergence of only one-third the magnitude of this difference is surprisingly small. Also, studies of primate behaviour are apparently not using exceedingly small sample sizes ($\tilde{x} = 16.5$) in comparison with research on other vertebrates, which may seem contradictory to common preconception.

Even if the general pattern matches the expectations, a closer look at the variation of sample sizes contained in the table does not reveal a standard procedure adopted in the choice of sample size. The observed substantial range cannot be explained by the criteria suggested to be useful in this decision process, indicating scope for improvement. We should be aware of the fact that choosing sample size is an optimization process. 'The more samples, the better' is not a sensible strategy, because it may waste resources and undermine ethical concerns. The sample size used in a study can be too small, which would compromise statistical power and might render the research effort useless, and it can be too large: 'while wasting time and energy on badly designed experiments is foolish, causing more human or animal suffering or more disturbance to an ecosystem than is absolutely necessary is inexcusable' (Ruxton & Colegrave 2006, p. 4). There may be cases where very large sample sizes are required because the effect size aimed to be identified is obscure or intentionally small, the random variation in underlying data cannot be reduced by prudent experimentation, or because there are important ethical, economical or societal reasons to minimize the Type-II error (β). Such causes typically apply in medical research, where, for instance, the intended effects or side-effects of a new drug are to be scrutinized. In basic research on behaviour, however, these conditions might be rare, and the temptation to inflate sample size for unjustified reasons should be countered (Still 1982).

Given there are reasons to minimize sample size, such as ethical considerations or other severe constraints (Allison et al. 1997), what is the minimum we should be content with? There is no easy answer to this question. One way is to apply sampling techniques allowing for a permanent adjustment of sample size during data acquisition (Armitage 1975). Such procedures using stopping rules are frequently used in clinical trials, but they are prone to Type-I error inflation and are limited to special conditions that often do not apply in behavioural research. Alternatively, a prudent decision can only be made by considering existing information, either from pilot experiments or previous studies, by using stringent experimentation and standardized conditions, and by a clear commitment to an approach that may somewhat compromise the Type-II error probability. This means that such study may not be publishable, because no effects can be detected by the appropriate analyses and it may seem probable that this was caused by insufficient statistical power. In other words, the results are inconclusive and few journals will devote precious publication space to their dissemination. On the flipside, should we publish studies based on small sample size if the latter turned out to be sufficient to reject the null hypothesis? One could argue that this increases the publication bias, which leads to an overrepresentation of 'positive' results in the scientific literature (Csada et al. 1996; Gurevitch & Hedges 1999). This may be a problem for meta-analyses that are based on published evidence (Song et al. 2000; Jennions & Møller 2002; Kotiaho & Tomkins 2002). However, there are strong arguments in favour of publishing such studies. Firstly, if authors can justify their choice of sample size satisfactorily, revealing that it resulted from prudent optimization instead of unwarranted austerity or neglect, there is no reason to conceal these results from the scientific public. Rather, such research might be commended for its rational proceeding. Secondly, it is unethical to withhold significant scientific results because this causes inevitable waste of human, animal and material resources. After all, these results may provide important information to subsequent studies regarding variance estimates and effect size (cf. Kelly 2006). And thirdly, refusing to publish such studies would open the floodgates to arbitrary censorship – who will take responsibility to decide whether a significant result was based on a 'sufficient' sample size, independently of the statistical procedure? And why then should studies not be rejected that were based on unjustifiably *large* samples as well?

If the policy must be to publish significant results even if based on small samples, what can be done about the problem that meta-analyses will be faced with a seemingly inevitable publication bias? Fortunately, there is evidence that a potential publication bias based on the detection of statistical significance (*sensu* Gurevitch & Hedges 1999) might not be as important as suspected (Jennions et al. 2004; Song et al. 2009) or even point in the opposite direction (in ecology: Koricheva 2003). In addition, there are ways to estimate the publication bias as long as information is provided in original research papers that allow to estimate effect size (Møller & Jennions 2001; Jennions & Møller 2002; Cassey et al. 2004), and there are procedures to reduce its influence in meta-analyses based on published studies (Lortie et al. 2007). A simple procedure is to confine the analysis to studies above a threshold sample size, which will reduce the number of publications that can be utilized but at the same time alleviate the problem. Mind that this is not a cause for not publishing studies based on small samples in the first place – after all, the value of a publication is only partly defined by its suitability for subsequent meta-analytical processing.

In conclusion, the choice of an adequate sample size in behavioural studies is neither trivial nor simple. Behavioural scientists do not appear to apply power analyses before collecting data but often seem to follow a 'best guess' strategy based on previous knowledge, considerations of the approach used and the costs and benefits involved in sampling. The use of experimental scrutiny as a means to reduce data variance seems to be of minor importance in the choice of sample size, which is difficult to understand. In any case, more transparency would be desirable about the selection of sample size in original studies. Preferably, this information should be given in the 'Methods' section of each article, for instance in a paragraph contained where the design of the study is described. Too small samples may entail inconclusive results, but it may be pointless to demand large sample sizes if ethical or other constraints make this prohibitive. On the other hand, both economy and ethical considerations should prevent us from undue inflation of sample size – optimization, not maximization should be the aim.

Appendix

Examples of free internet sources for power and sample size calculations:

http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register

http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize

http://statpages.org/#Power
http://www.stat.uiowa.edu/~rlenth/Power/index.html
http://www.epibiostat.ucsf.edu/biostat/sampsize.html

Literature Cited

Allison, D. B., Allison, R. L., Faith, M. S., Paultre, F. & Pi-Sunyer, F. X. 1997: Power and money: designing statistically powerful studies while minimizing financial costs. Psychol. Methods **2**, 20—33.

Anderson, D. R., Burnham, K. P. & Thompson, W. L. 2000: Null hypothesis testing: problems, prevalence, and an alternative. J. Wildl. Manage. **64**, 912—923.

Anderson, D. R., Link, W. A., Johnson, D. H. & Burnham, K. P. 2001: Suggestions for presenting the results of data analyses. J. Wildl. Manage. **65**, 373—378.

Armitage, P. 1975: Sequential Medical Trials. Blackwell, Oxford.

Baguley, T. 2004: Understanding statistical power in the context of applied research. Appl. Ergon. **35**, 73—80.

Ballentine, B. 2009: The ability to perform physically challenging songs predicts age and size in male swamp sparrows, *Melospiza georgiana*. Anim. Behav. **77**, 973—978.

Balsby, T. J. S. & Bradbury, J. W. 2009: Vocal matching by orange-fronted conures (*Aratinga canicularis*). Behav. Processes **82**, 133—139.

Beaulieu, M., Thierry, A. M., Raclot, T., Maho, Y., Ropert-Coudert, Y., Gachot-Neveu, H. & Ancel, A. 2009: Sex-specific parental strategies according to the sex of offspring in the Adelie penguin. Behav. Ecol. **20**, 878—883.

Benedict, L. & McEntee, J. P. 2009: Context, structural variability and distinctiveness of California towhee (*Pipilo crissalis*) vocal duets. Ethology **115**, 77—86.

Bijleveld, A. I. & Mullers, R. H. E. 2009: Reproductive effort in biparental care: an experimental study in long-lived Cape gannets. Behav. Ecol. **20**, 736—744.

Bissonnette, A., Lange, E. & van Schaik, C. P. 2009: A cardinal measure of competitive ability in Barbary macaque males (*Macaca sylvanus*). Ethology **115**, 671—681.

Bourjade, M., Thierry, B., Maumy, M. & Petit, O. 2009: Decision-making in Przewalski horses (*Equus ferus przewalskii*) is driven by the ecological contexts of collective movements. Ethology **115**, 321—330.

Brauer, M. & McClelland, G. 2005: The use of contrasts in data analysis: how to test specific hypotheses in psychological research. Annee Psychol. **105**, 273—305.

Burnham, K. P. & Anderson, D. R. 2002: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag, Berlin.

Busquet, N., Nizerolle, C. L. & Feron, C. 2009: What triggers reproductive life? Effects of adolescent cohabitation, social novelty and aggression in a monogamous mouse. Ethology **115**, 87—95.

Campbell, D. L. M., Shaw, R. C. & Hauber, M. E. 2009: The strength of species recognition in captive female zebra finches (*Taeniopygia guttata*): a comparison across Estrildid heterospecifics. Ethology **115**, 23—32.

Cardoso, G. C., Atwell, J. W., Ketterson, E. D. & Price, T. D. 2009: Song types, song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). Behav. Ecol. **20**, 901—907.

Carter, A. J., Macdonald, S. L., Thomson, V. A. & Goldizen, A. W. 2009: Structured association patterns and their energetic benefits in female eastern grey kangaroos, *Macropus giganteus*. Anim. Behav. **77**, 839—846.

Cassey, P., Ewen, J. G., Blackburn, T. M. & Møller, A. P. 2004: A survey of publication bias within evolutionary ecology. Proc. R. Soc. Lond. B Biol. Sci. **271**, S451—S454.

ten Cate, C., Bruins, W. S., den Ouden, J., Egberts, T., Neevel, H., Spierings, M., van der Burg, K. & Brokerhof, A. W. 2009: Tinbergen revisited: a replication and extension of experiments on the beak colour preferences of herring gull chicks. Anim. Behav. **77**, 795—802.

Chapman, N. C., Nanork, P., Gloag, R. S., Wattanachaiyingcharoen, W., Beekman, M. & Oldroyd, B. P. 2009: Queenless colonies of the Asian red dwarf honey bee (*Apis florea*) are infiltrated by workers from other queenless colonies. Behav. Ecol. **20**, 817—820.

Chelini, M. C., Willemart, R. H. & Hebets, E. A. 2009: Costs and benefits of freezing behaviour in the harvestman *Eumesosoma roeweri* (Arachnida, Opiliones). Behav. Processes **82**, 153—159.

Cohen, J. 1962: The statistical power of abnormal-social psychological research: a review. J. Abnorm. Soc. Psychol. **65**, 145—153.

Cohen, J. 1988: Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale, NJ.

Colegrave, N. & Ruxton, G. D. 2003: Confidence intervals are a more useful complement to nonsignificant tests than are power calculations. Behav. Ecol. **14**, 446—447.

Cooper, W. E., Wilson, D. S. & Smith, G. R. 2009a: Sex, reproductive status, and cost of tail autotomy via decreased running speed in lizards. Ethology **115**, 7—13.

Cooper, W. E., Caldwell, J. P. & Vitt, L. J. 2009b: Risk assessment and withdrawal behavior by two species of aposematic poison frogs, *Dendrobates auratus* and *Oophaga pumilio*, on forest trails. Ethology **115**, 311—320.

Costello, A. K., Pultorak, J. D. & Meikle, D. B. 2009: Do male house mice (*Mus musculus*) discriminate between females that differ in nutritional status? Behav. Processes **82**, 119—125.

Csada, R. D., James, P. C. & Espie, R. H. M. 1996: The ''file drawer problem'' of non-significant results: does it apply to biological research? Oikos **76**, 591—593.

Dickinson, J. L., Euaparadorn, M., Greenwald, K., Mitra, C. & Shizuka, D. 2009: Cooperation and competition: nepotistic tolerance and intrasexual aggression in western bluebird winter groups. Anim. Behav. **77**, 867—872.

Donohoe, M. L., von Hippel, W. & Brooks, R. C. 2009: Beyond waist-hip ratio: experimental multivariate evidence that average women's torsos are most attractive. Behav. Ecol. **20**, 716—721.

Dupont, W. D. & Plummer, W. D. 1998: Power and sample size calculations for studies involving linear regression. Control. Clin. Trials **19**, 589—601.

Eraly, D., Hendrickx, F. & Lens, L. 2009: Condition-dependent mate choice and its implications for population differentiation in the wolf spider *Pirata piraticus*. Behav. Ecol. **20**, 856—863.

Fan, C. M., Yang, E. C. & Tso, I. M. 2009: Hunting efficiency and predation risk shapes the color-associated foraging traits of a predator. Behav. Ecol. **20**, 808—816.

Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. 2007: G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods **39**, 175—191.

Felton, A. M., Felton, A., Raubenheimer, D., Simpson, S. J., Foley, W. J., Wood, J. T., Wallis, I. R. & Lindenmayer, D. B. 2009: Protein content of diets dictates the daily energy intake of a free-ranging primate. Behav. Ecol. **20**, 685—690.

Fernandez-Montraveta, C. & Cuadrado, M. 2009: Mate attraction in a burrowing wolf-spider (Araneae, Lycosidae) is not olfactory mediated. Ethology **115**, 375—383.

Freidin, E., Cuello, M. I. & Kacelnik, A. 2009: Successive negative contrast in a bird: starlings' behaviour after unpredictable negative changes in food quality. Anim. Behav. **77**, 857—865.

Fucikova, E., Drent, P. J., Smits, N. & van Oers, K. 2009: Handling stress as a measurement of personality in great tit nestlings (*Parus major*). Ethology **115**, 366—374.

Fureix, C., Pages, M., Bon, R., Lassalle, J. M., Kuntz, P. & Gonzalez, G. 2009: A preliminary study of the effects of handling type on horses' emotional reactivity and the human–horse relationship. Behav. Processes **82**, 202—210.

Galeotti, P., Rubolini, D., Pupin, F., Sacchi, R., Altobelli, E., Nardi, P. A. & Fasola, M. 2009: Presence of rivals reduces mating probability but does not affect ejaculate size in the freshwater crayfish *Austropotamobius italicus*. Behaviour **146**, 45—68.

Garamszegi, L. Z., Eens, M. & Janos, T. 2009: Behavioural syndromes and trappability in free-living collared flycatchers, *Ficedula albicollis*. Anim. Behav. **77**, 803—812.

Gero, S., Engelhaupt, D., Rendell, L. & Whitehead, H. 2009: Who cares? Between-group variation in alloparental caregiving in sperm whales. Behav. Ecol. **20**, 838—843.

Gilbert-Norton, L. B., Shahan, T. A. & Shivik, J. A. 2009: Coyotes (*Canis latrans*) and the matching law. Behav. Processes **82**, 178—183.

Gipson, C. D., Miller, H. C., Alessandri, J. J. D. & Zentall, T. R. 2009: Within-trial contrast: the effect of probability of reinforcement in training. Behav. Processes **82**, 126—132.

Green, K. & Tregenza, T. 2009: The influence of male ejaculates on female mate search behaviour, oviposition and longevity in crickets. Anim. Behav. **77**, 887—892.

Griesser, M. & Nystrand, M. 2009: Vigilance and predation of a forest-living bird species depend on large-scale habitat structure. Behav. Ecol. **20**, 709—715.

Grim, T., Rutila, J., Cassey, P. & Hauber, M. E. 2009: Experimentally constrained virulence is costly for common cuckoo chicks. Ethology **115**, 14—22.

Gurevitch, J. & Hedges, L. V. 1999: Statistical issues in ecological meta-analyses. Ecology **80**, 1142—1149.

Hansen, H., McDonald, D. B., Groves, P., Maier, J. A. K. & Ben David, M. 2009: Social networks and the formation and maintenance of river otter groups. Ethology **115**, 384—396.

Harvey, D. R., Brant, L. & Commins, S. 2009: Differences in cue-dependent spatial navigation may be revealed by in-depth swimming analysis. Behav. Processes **82**, 190—197.

Hayworth, M. K., Johnson, N. G., Wilhelm, M. E., Gove, R. P., Metheny, J. D. & Rueppell, O. 2009: Added weights lead to reduced flight behavior and mating success in polyandrous honey bee queens (*Apis mellifera*). Ethology **115**, 698—706.

Hobbs, N. T. & Hilborn, R. 2006: Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecol. Appl. **16**, 5—19.

Hodge, S. J., Bell, M. B. V., Mwanguhya, F., Kyabulima, S., Waldick, R. C. & Russell, A. F. 2009: Maternal weight, offspring competitive ability, and the evolution of communal breeding. Behav. Ecol. **20**, 729—735.

Hoefler, C. D., Guhanarayan, G., Persons, M. H. & Rypstra, A. L. 2009: The interaction of female condition and mating status on male–male aggression in a wolf spider. Ethology **115**, 331—338.

Hoenig, J. M. & Heisey, D. M. 2001: The abuse of power: the pervasive fallacy of power calculations for data analysis. Am. Stat. **55**, 19—24.

Hohberg, K. & Traunspurger, W. 2009: Foraging theory and partial consumption in a tardigrade-nematode system. Behav. Ecol. **20**, 884—890.

Holbrook, C. T., Clark, R. M., Jeanson, R., Bertram, S. M., Kukuk, P. F. & Fewell, J. H. 2009: Emergence and consequences of division of labor in associations of normally solitary sweat bees. Ethology **115**, 301—310.

Howard, G. S., Maxwell, S. E. & Fleming, K. J. 2000: The proof of the pudding: an illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. Psychol. Methods **5**, 315—332.

Jacob, A., Evanno, G., Renai, E., Sermier, R. & Wedekind, C. 2009: Male body size and breeding tubercles are both linked to intrasexual dominance and reproductive success in the minnow. Anim. Behav. **77**, 823—829.

Jennions, M. D. & Møller, A. P. 2002: Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. Biol. Rev. **77**, 211—222.

Jennions, M. D. & Møller, A. P. 2003: A survey of the statistical power of research in behavioral ecology and animal behavior. Behav. Ecol. **14**, 438—445.

Jennions, M. D., Møller, A. P. & Hunt, J. 2004: Meta-analysis can ''fail'': reply to Kotiaho and Tomkins. Oikos **104**, 191—193.

Johnson, D. H. 1999: The insignificance of statistical significance testing. J. Wildl. Manage. **63**, 763—772.

Kangas, B. D., Berry, M. S., Cassidy, R. N., Dallery, J., Vaidya, M. & Hackenberg, T. D. 2009: Concurrent performance in a three-alternative choice situation: response allocation in a rock/paper/scissors game. Behav. Processes **82**, 164—172.

Karino, K. & Sato, A. 2009: Male-biased sex ratios in offspring of attractive males in the guppy. Ethology **115**, 682—690.

Kelly, C. D. 2006: Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. Q. Rev. Biol. **81**, 221—236.

Kojima, W., Kitamura, W., Kitajima, S., Ito, Y., Ueda, K., Fujita, G. & Higuchi, H. 2009: Female barn swallows gain indirect but not direct benefits through social mate choice. Ethology **115**, 939—947.

Kontiainen, P., Pietiainen, H., Huttunen, K., Karell, P., Kolunen, H. & Brommer, J. E. 2009: Aggressive Ural owl mothers recruit more offspring. Behav. Ecol. **20**, 789—796.

Koricheva, J. 2003: Non-significant results in ecology: a burden or a blessing in disguise? Oikos **102**, 397—401.

Kotiaho, J. S. & Tomkins, J. L. 2002: Meta-analysis, can it ever fail? Oikos **96**, 551—553.

Kraemer, H. C. & Thiemann, S. 1987: How Many Subjects? Sage Publications, London.

Langridge, K. V. 2009: Cuttlefish use startle displays, but not against large predators. Anim. Behav. **77**, 847—856.

Lenth, R. V. 2001: Some practical guidelines for effective sample size determination. Am. Stat. **55**, 187—193.

Lewis, K. P. 2006: Statistical power, sample sizes, and the software to calculate them easily. Bioscience **56**, 607—612.

Lewis, S. M. & Rotjan, R. D. 2009: Vacancy chains provide aggregate benefits to *Coenobita clypeatus* hermit crabs. Ethology **115**, 356—365.

Lindqvist, C., Lind, J. & Jensen, P. 2009: Effects of domestication on food deprivation-induced behaviour in red junglefowl, *Gallus gallus*, and White Leghorn layers. Anim. Behav. **77**, 893—899.

Lisonbee, L. D., Villalba, J. J., Provenza, F. D. & Hall, J. O. 2009: Tannins and self-medication: implications for sustainable parasite control in herbivores. Behav. Processes **82**, 184—189.

Logue, D. M., Mishra, S., McCaffrey, D., Ball, D. & Cade, W. H. 2009: A behavioral syndrome linking courtship behavior toward males and females predicts reproductive success from a single mating in the hissing cockroach, *Gromphadorhina portentosa*. Behav. Ecol. **20**, 781—788.

Lohrey, A. K., Clark, D. L., Gordon, S. D. & Uetz, G. W. 2009: Antipredator responses of wolf spiders (Araneae: Lycosidae) to sensory cues representing an avian predator. Anim. Behav. **77**, 813—821.

Lomborg, J. P. & Toft, S. 2009: Nutritional enrichment increases courtship intensity and improves mating success in male spiders. Behav. Ecol. **20**, 700—708.

Lortie, C. J., Aarssen, L. W., Budden, A. E., Koricheva, J. K., Leimu, R. & Tregenza, T. 2007: Publication bias and merit in ecology. Oikos **116**, 1247—1253.

Loveless, S. A., Kittell, M. M. & Ptacek, M. B. 2009: Patterns of inheritance of mating signals in sailfin and shortfin mollies (Poeciliidae: Poecilia: Mollienesia). Ethology **115**, 958—971.

Magrath, R. D., Pitcher, B. J. & Gardner, J. L. 2009: An avian eavesdropping network: alarm signal reliability and heterospecific response. Behav. Ecol. **20**, 745—752.

Majolo, B., Ventura, R. & Koyama, N. F. 2009a: Anxiety level predicts post-conflict behaviour in wild Japanese macaques (*Macaca fuscata yakui*). Ethology **115**, 986—995.

Majolo, B., Ventura, R., Koyama, N. F., Hardie, S. M., Jones, B. M., Knapp, L. A. & Schino, G. 2009b: Analysing the effects of group size and food competition on Japanese macaque social relationships. Behaviour **146**, 113—137.

Maresova, J., Krasa, A. & Frynta, D. 2009: We all appreciate the same animals: cross-cultural comparison of human aesthetic preferences for snake species in Papua New Guinea and Europe. Ethology **115**, 297—300.

Martinez, L. & Matsuzawa, T. 2009: Effect of species-specificity in auditory-visual intermodal matching in a chimpanzee (*Pan troglodytes*) and humans. Behav. Processes **82**, 160—163.

Maxwell, S. E. 2004: The persistence of underpowered studies in psychological research: causes, consequences, and remedies. Psychol. Methods **9**, 147—163.

Maxwell, S. E., Kelley, K. & Rausch, J. R. 2008: Sample size planning for statistical power and accuracy in parameter estimation. Annu. Rev. Psychol. **59**, 537—563.

McCarthy, M. A. & Masters, P. 2005: Profiting from prior information in Bayesian analyses of ecological data. J. Appl. Ecol. **42**, 1012—1019.

McClelland, G. H. 1997: Optimal design in psychological research. Psychol. Methods **2**, 3—19.

McClelland, G. H. 2000: Increasing statistical power without increasing sample size. Am. Psychol. **55**, 963—964.

McDonald, P. G., Wilson, D. R. & Evans, C. S. 2009: Nestling begging increases predation risk, regardless of spectral characteristics or avian mobbing. Behav. Ecol. **20**, 821—829.

McLean, I. G., Cameron, E. Z., Linklater, W. L., Schmitt, N. T. & Pulskamp, K. S. M. 2009: Partnerships in the social system of a small macropod marsupial, the quokka (*Setonix brachyurus*). Behaviour **146**, 89—112.

McNamara, K. B., Elgar, M. A. & Jones, T. M. 2009: Large spermatophores reduce female receptivity and increase male paternity success in the almond moth, *Cadra cautella*. Anim. Behav. **77**, 931—936.

Meldrum, G. E. & Ruckstuhl, K. E. 2009: Mixed-sex group formation by bighorn sheep in winter: trading costs of synchrony for benefits of group living. Anim. Behav. **77**, 919—929.

Minderman, J., Reid, J. M., Evans, P. G. H. & Whittingham, M. J. 2009: Personality traits in wild starlings: exploration behavior and environmental sensitivity. Behav. Ecol. **20**, 830—837.

Møller, A. P. & Jennions, M. D. 2001: Testing and adjusting for publication bias. Trends Ecol. Evol. **16**, 580—586.

Møller, A. P. & Jennions, M. D. 2002: How much variance can be explained by ecologists and evolutionary biologists? Oecologia **132**, 492—500.

Molnár, C., Pongracz, P., Farago, T., Doka, A. & Miklosi, A. 2009: Dogs discriminate between barks: the effect of context and identity of the caller. Behav. Processes **82**, 198—201.

Morand-Ferron, J., Lalande, E. & Giraldeau, L. A. 2009: Large-scale input matching by urban feral pigeons (*Columba livia*). Ethology **115**, 707—712.

Muscedere, M. L., Willey, T. A. & Traniello, J. F. A. 2009: Age and task efficiency in the ant *Pheidole dentata*: young minor workers are not specialist nurses. Anim. Behav. **77**, 911—918.

Nakagawa, S. 2004: A farewell to Bonferroni: the problems of low statistical power and publication bias. Behav. Ecol. **15**, 1044—1045.

Nakagawa, S. & Cuthill, I. C. 2007: Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol. Rev. **82**, 591—605.

Nakagawa, S. & Foster, T. M. 2004: The case against retrospective statistical power analyses with an introduction to power analysis. Acta Ethol. **7**, 103—108.

Nessler, S. H., Uhl, G. & Schneider, J. M. 2009: Scent of a woman – the effect of female presence on sexual cannibalism in an orb-weaving spider (Araneae: Araneidae). Ethology **115**, 633—640.

Nilsson, P. A., Jacobsen, L., Berg, S. & Skov, C. 2009: Environmental conditions and intraspecific interference: unexpected effects of turbidity on pike (*Esox lucius*) foraging. Ethology **115**, 33—38.

Orrock, J. L. & Danielson, B. J. 2009: Temperature and cloud cover, but not predator urine, affect winter foraging of mice. Ethology **115**, 641—648.

Padilla, D. P. & Nogales, M. 2009: Behavior of kestrels feeding on frugivorous lizards: implications for secondary seed dispersal. Behav. Ecol. **20**, 872—877.

Palacin, C., Alonso, J. C., Alonso, J. A., Martin, C. A., Magana, M. & Martin, B. 2009: Differential migration by sex in the great bustard: possible consequences of an extreme sexual size dimorphism. Ethology **115**, 617—626.

Pareja, M., Mohib, A., Birkett, M. A., Dufour, S. & Glinwood, R. T. 2009: Multivariate statistics coupled to generalized linear models reveal complex use of chemical cues by a parasitoid. Anim. Behav. **77**, 901—909.

Parsons, K. M., Balcomb, K. C., Ford, J. K. B. & Durban, J. W. 2009: The social dynamics of southern resident killer whales and conservation implications for this endangered population. Anim. Behav. **77**, 963—971.

Patzelt, A., Pirow, R. & Fischer, J. 2009: Post-conflict affiliation in Barbary macaques is influenced by conflict characteristics and relationship quality, but does not diminish short-term renewed aggression. Ethology **115**, 658—670.

Petter, M., Musolino, E., Roberts, W. A. & Cole, M. 2009: Can dogs (*Canis familiaris*) detect human deception? Behav. Processes **82**, 109—118.

Pillay, N. & Kinahan, A. A. 2009: Mating strategy predicts the occurrence of the Bruce effect in the vlei rat *Otomys irroratus*. Behaviour **146**, 139—151.

Polak, M. & Simmons, L. W. 2009: Secondary sexual trait size reveals competitive fertilization success in *Drosophila bipectinata* Duda. Behav. Ecol. **20**, 753—760.

Quinn, G. P. & Keough, M. J. 2002: Experimental Design and Data Analysis for Biologists. Cambridge Univ. Press, Cambridge.

Ramseyer, A., Thierry, B., Boissy, A. & Dumont, B. 2009: Decision-making processes in group departures of cattle. Ethology **115**, 948—957.

Rao, D. 2009: Experimental evidence for the amelioration of shadow competition in an orb-web spider through the 'Ricochet' effect. Ethology **115**, 691—697.

Ratterman, N. L., Rosenthal, G. G. & Jones, A. G. 2009: Sex recognition via chemical cues in the sex-role-reversed gulf pipefish (*Syngnathus scovelli*). Ethology **115**, 339—346.

Robbins, A. M., Stoinski, T., Fawcett, K. & Robbins, M. M. 2009: Leave or conceive: natal dispersal and philopatry of female mountain gorillas in the Virunga volcano region. Anim. Behav. **77**, 831—838.

Rosnow, R. L. & Rosenthal, R. 2003: Effect sizes for experimenting psychologists. Can. J. Exp. Psychol. **57**, 221—237.

Ruxton, G. D. & Beauchamp, G. 2008: Time for some a priori thinking about post hoc testing. Behav. Ecol. **19**, 690—693.

Ruxton, G. D. & Colegrave, N. 2006: Experimental Design for the Life Sciences, 2nd edn. Oxford Univ. Press, Oxford.

Rypstra, A. L., Schlosser, A. M., Sutton, P. L. & Persons, M. H. 2009: Multimodal signalling: the relative importance of chemical and visual cues from females to the behaviour of male wolf spiders (Lycosidae). Anim. Behav. **77**, 937—947.

Salomon, M. 2009: Social environment and feeding state influence movement decisions in a web-building spider. Ethology **115**, 916—927.

Schino, G., Di Giuseppe, F. & Visalberghi, E. 2009: The time frame of partner choice in the grooming reciprocation of *Cebus apella*. Ethology **115**, 70—76.

Schroeder, J., Lourenco, P. M., Hooijmeijer, J. C. E. W., Both, C. & Piersma, T. 2009: A possible case of contemporary selection leading to a decrease in sexual plumage dimorphism in a grassland-breeding shorebird. Behav. Ecol. **20**, 797—807.

Schubert, M., Pillay, N., Ribble, D. O. & Schradin, C. 2009: The round-eared sengi and the evolution of social monogamy: factors that constrain males to live with a single female. Ethology **115**, 972—985.

Schwagmeyer, P. L. & Mock, D. W. 1997: How to minimize sample sizes while preserving statistical power. Anim. Behav. **54**, 470—474.

Segoli, M., Harari, A. R., Bouskila, A. & Keasar, T. 2009: Brood size in a polyembryonic parasitoid wasp is affected by relatedness among competing larvae. Behav. Ecol. **20**, 761—767.

Silva, K., Almada, V. C., Vieira, M. N. & Monteiro, N. M. 2009: Female reproductive tactics in a sex-role reversed pipefish: scanning for male quality and number. Behav. Ecol. **20**, 768—772.

Song, F., Eastwood, A., Gilbody, S., Duley, L. & Sutton, A. 2000: Publication and related biases. Health Technol. Assess. **4**, 1—115.

Song, F. J., Parekh-Bhurke, S., Hooper, L., Loke, Y. K., Ryder, J. J., Sutton, A. J., Hing, C. B. & Harvey, I. 2009: Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. BMC Med. Res. Methodol. **9**, 79.

Steidl, R. J. & Thomas, L. 2001: Power analysis and experimental design. In: Design and Analysis of Ecological Experiments, 2nd edn (Scheiner, S. M. & Gurevitch, J., eds). Oxford Univ. Press, Oxford, pp. 14—36.

Stephens, P. A., Buskirk, S. W. & del Rio, C. M. 2007: Inference in ecology and evolution. Trends Ecol. Evol. **22**, 192—197.

Still, A. W. 1982: On the number of subjects used in animal behaviour experiments. Anim. Behav. **30**, 873—880.

Strausberger, B. M. & Rothstein, S. I. 2009: Parasitic cowbirds may defeat host defense by causing rejecters to misimprint on cowbird eggs. Behav. Ecol. **20**, 691—699.

Suzuki, S. & Nagano, M. 2009: To compensate or not? Caring parents respond differentially to mate removal and mate handicapping in the burying beetle, *Nicrophorus quadripunctatus*. Ethology **115**, 1—6.

Takahashi, M. K. & Parris, M. J. 2009: When do the costs of spermatogenesis constrain sperm expenditure? Remarks on the pattern of the spermatogenic cycle. Ethology **115**, 96—104.

Taylor, A. M., Reby, D. & McComb, K. 2009: Context-related variation in the vocal growling behaviour of the domestic dog (*Canis familiaris*). Ethology **115**, 905—915.

Tinghitella, R. M., Wang, J. M. & Zuk, M. 2009: Preexisting behavior renders a mutation adaptive: flexibility in male phonotaxis behavior and the loss of singing ability in the field cricket *Teleogryllus oceanicus*. Behav. Ecol. **20**, 722—728.

Tóth, Z., Bokony, V., Lendvai, A. Z., Szabo, K., Penzes, Z. & Liker, A. 2009: Whom do the sparrows follow? The effect of kinship on social preference in house sparrow flocks. Behav. Processes **82**, 173—177.

Trefry, S. A. & Hik, D. S. 2009: Eavesdropping on the neighbourhood: collared pika (*Ochotona collaris*) responses to playback calls of conspecifics and heterospecifics. Ethology **115**, 928—938.

Truppa, V., Spinozzi, G., Stegagno, T. & Fagot, J. 2009: Picture processing in tufted capuchin monkeys (*Cebus apella*). Behav. Processes **82**, 140—152.

VanderWaal, K. L., Mosser, A. & Packer, C. 2009: Optimal group size, dispersal decisions and postdispersal relationships in female African lions. Anim. Behav. **77**, 949—954.

Verdolin, J. L. & Slobodchikoff, C. N. 2009: Resources, not kinship, determine social patterning in the territorial Gunnison's prairie dog (*Cynomys gunnisoni*). Ethology **115**, 59—69.

Verzijden, M. N., Zwinkels, J. & ten Cate, C. 2009: Cross-fostering does not influence the mate preferences and territorial behaviour of males in Lake Victoria cichlids. Ethology **115**, 39—48.

Wascher, C. A. F., Scheiber, I. B. R., Weiss, B. M. & Kotrschal, K. 2009: Heart rate responses to agonistic encounters in greylag geese, *Anser anser*. Anim. Behav. **77**, 955—961.

Wiebe, K. L. & Slagsvold, T. 2009: Parental sex differences in food allocation to junior brood members as mediated by prey size. Ethology **115**, 49—58.

Wilgers, D. J., Nicholas, A. C., Reed, D. H., Stratton, G. E. & Hebets, E. A. 2009: Condition-dependent alternative mating tactics in a sexually cannibalistic wolf spider. Behav. Ecol. **20**, 891—900.

Wilson, D. R., Nelson, X. J. & Evans, C. S. 2009: Seizing the opportunity: subordinate male fowl respond rapidly to variation in social context. Ethology **115**, 996—1004.

Wroblewski, E. E., Murray, C. M., Keele, B. F., Schumacher-Stankey, J. C., Hahn, B. H. & Pusey, A. E. 2009: Male dominance rank and reproductive success in chimpanzees, *Pan troglodytes schweinfurthii*. Anim. Behav. **77**, 873—885.

Zaatari, D., Palestis, B. G. & Trivers, R. 2009: Fluctuating asymmetry of responders affects offers in the ultimatum game oppositely according to attractiveness or need as perceived by proposers. Ethology **115**, 627—632.

Zanette, L. & Field, J. 2009: Cues, concessions, and inheritance: dominance hierarchies in the paper wasp *Polistes dominulus*. Behav. Ecol. **20**, 773—780.