# Exaggerated sexual swellings in female non-human primates are reliable signals of female fertility and body condition

Sally E. Street[1,2,3], Catharine P. Cross[1] and Gillian R. Brown[1*]

[1]School of Psychology & Neuroscience, University of St Andrews, UK
[2]School of Biology, University of St Andrews, UK
[3]School of Biological, Biomedical and Environmental Sciences, University of Hull, UK

In press: *Animal Behaviour*

* Correspondence: G. R. Brown, School of Psychology & Neuroscience, University of St Andrews, South Street, St Andrews, Fife, KY16 8NF, UK
*Email address*: grb4@st-andrews.ac.uk (G. R. Brown)

## Highlights

- In some primates, females exhibit exaggerated swellings of the anogenital region.
- Maximally swollen females receive the highest levels of mating interest from males.
- Our meta-analysis shows that swelling size is closely related to fertility.
- We show that swelling size is also positively correlated with body condition.
- The results support both the graded-signal and reliable-indicator hypotheses.

## Abstract

In some species of Old World monkeys and apes, females exhibit exaggerated swellings of the anogenital region that vary in size across the ovarian cycle. Exaggerated swellings are typically largest around the time of ovulation, and swelling size has been reported to correlate positively with female quality, supporting the hypothesis that exaggerated swellings are honest signals of both female fecundity and quality. However, the relationship between swelling size and timing of ovulation is weak in some studies, and the relationship between swelling size and female quality has also not been consistently reported. Here, we collated empirical studies that have reported either i) swelling size and estimated timing of ovulation (*N*=26), or ii) swelling size and measures of individual quality (*N*=7), to assess the strength of these relationships using meta-analytic methods. Our analyses confirmed that the period of maximal swelling size is closely associated with the most fertile period of the ovarian cycle and that a large proportion of ovulations occur during the maximal swelling period. A small, positive effect size was also found for the relationship between swelling size and body condition. In contrast, the relationships with age and social rank were not significant. Swelling size, therefore, potentially signals both female condition and timing of the fertile phase. Males are likely to benefit from allocating mating effort according to swelling size, while females with large swellings potentially benefit from exerting control over matings in species where female control is compromised by male mating strategies.

Key words: ornament, graded signal, ovulation, reliable indicator, quality

## Introduction

Exaggerated swellings of the female anogenital region occur in around 30 species of Old World monkeys and apes (Sillén-Tullberg and Møller, 1993; Dixson, 2012). These swellings have been described as female ornaments (Clutton-Brock, 2007; 2009), defined as elaborate traits that function to attract males. The tissues of the anogenital region swell as a result of water retention and become red in colouration through distention of the capillaries (Dixson, 2012). Experimental studies have shown that swelling size and colouration increases in response to oestrogen during the follicular phase of the ovarian cycle (Dixson, 2012). The tissue then rapidly detumesces and returns to normal colouration after ovulation (e.g., Brauch et al., 2007). Females potentially benefit from displaying swellings through attracting a preferred mating partner and/or multiple partners, while males potentially benefit from strategically allocating mating effort in relation to within- and between-female variation in fertility status (Nunn, 1999; Pagel, 1994).

Observational studies have shown that maximally swollen females receive the highest levels of mating interest from males (e.g., Gesquiere et al., 2007), and experimental studies have demonstrated that male baboons are sexually aroused by artificial swellings (Bielert and Anderson, 1985; Girolami and Bielert, 1987). Females are also most likely to approach males and solicit copulations when maximally swollen (Higham et al., 2012). In most species with exaggerated swellings, females mate with more than one male during an ovarian cycle (Clutton-Brock and Harvey, 1976; Pagel and Meade, 2006), and low-ranking males tend to mate earlier in the swelling period than high-ranking males (Gesquiere et al., 2007; Emery Thompson, 2005). These observations are consistent with exaggerated swellings functioning as ‘graded’ signals of the probability of ovulation that provide females with the benefits of mating with multiple males, such as protection from infanticide, while also biasing paternity towards high-quality males (Nunn, 1999; Alberts & Fitzpatrick, 2012).

Researchers have also suggested that exaggerated swellings are relatively unreliable signals of fertility at least in some species (van Schaik et al., 2000; Zinner et al., 2004), given that swellings are often reported to reach a large size several days before ovulation (Deschner et al., 2003; Reichert et al., 2002). For example, a study of bonobos (*Pan paniscus*) reported that females were maximally swollen for an average of 16 days (Reichert et al., 2002), while the fertile window is likely to last around 4 days (Dunson et al., 1999; Wilcox et al., 1995). The same study reported that 30 percent of ovulations occurred outside of the maximal swelling period (Reichert et al., 2002), further supporting the idea that swellings are an unreliable signal of fertility in this species. Selection has thus been suggested to favour mechanisms that make ovulation unpredictable relative to swelling characteristics (van Schaik et al., 2000), with the null hypothesis being that some error in signalling occurs by chance.

Exaggerated swellings have also been reported to vary in maximal size both between individuals and across cycles within individuals (Fitzpatrick et al., 2014). For instance, maximal swelling size increases with the number of cycles since resumption of cycling following a previous birth (Higham et al., 2008; Huchard et al., 2009). Exaggerated swellings are costly in terms of increased body weight and risk of tissue damage (Bielert and Busse, 1983; Matsumodo-Oda, 1998), and females receive greater amounts of physical harassment from both male and female group members when maximally swollen (Huchard and Cowlishaw, 2011, Feldblum et al., 2014). Thus, swelling size is potentially indicative of a female’s current physical condition and ability to raise an offspring (Pagel, 1994). While this ‘reliable indicator’ hypothesis has received some support (Huchard et al., 2009; Domb and Pagel, 2001), other studies failed to find a correlation between swelling size and measures of female quality (Möhle et al., 2005; Setchell and Wickings, 2004), where quality is broadly defined as any property of an individual that influences the cost of expressing the exaggerated trait (Johnstone et al., 2009).

Using currently available data, we first investigated the strength of the evidence that exaggerated swellings are generally reliable signals that peak in size during the period of highest fertility. Given that reliability is defined as the tightness of the correlation with the underlying trait (Searcy and Nowicki, 2005), we predicted that the period of maximal swelling would not, on average, extend greatly beyond the fertile period and that a relatively strong association would be found between swelling size and the probability of ovulation. Because the strength of the association between swelling size and fertility is likely to depend upon how swelling size is measured, we also compared studies that used fine-resolution, continuous measures of swelling size with those that used broad-resolution, categorical measures, with the prediction that studies using categorical measures could under-estimate the precision of the relationship between swelling size and ovulation probability. Finally, we conducted meta-analyses on studies that have correlated swelling size with measures of female properties other than current fertility status to test whether exaggerated swellings are reliable indicators of female quality.

## Methods

### *Literature review and inclusion criteria*

Articles were located using keyword searches in Thomson Reuter's *Web of Knowledge* (e.g., 'primate swelling ovulation', 'primate swelling quality') and by checking the reference sections of papers that were identified. To fulfil the inclusion criteria, articles had to contain either a) data on changes in swelling size across days relative to an independent estimate of the day of ovulation based on ovarian hormone levels (from blood, urine or fecal samples) or laparoscopy, or b) data on swelling size and any measures of individual female quality (e.g., body condition or social rank). The final dataset on swelling size and ovulation consisted of 26 articles (including Deschner et al., 2004, which reported swelling size from the same subjects as Deschner et al., 2003 using a different measure) across 10 species (*Cercocebus atys, Macaca nemestrina,* *Macaca sylvanus, Macaca nigra*, *Macaca tonkeana*, *Mandrillus sphinx*, *Pan paniscus*, *Pan troglodytes*, *Papio anubis*, *Papio cynocephalus*) (Table A1 in Appendix). A small number of studies were excluded, as data on the day of ovulation in relation to maximal swelling size could not be extracted (i.e., Bullock et al., 1972; Dahl et al., 1991; Fürtbauer et al., 2010; Nyakudya et al., 2012; Shaikh et al., 1982; Thomson et al., 1992; Wildt et al., 1977). As exaggerated swellings are the focus of existing functional hypotheses (e.g., Pagel, 1994; Nunn, 1999), data from species with small swellings (defined as swellings that involve the vulval and clitoral areas only; Dixson, 2012) were excluded (i.e., *Hylobates lar*: Barelli et al., 2007; Nadler et al., 1993; *Gorilla gorilla*: Czekala and Sicotte, 2000; Nadler et al., 1979), as were data from species with sub-caudal swellings (i.e., *Macaca fascicularis*: Engelhardt et al., 2005). The relationship between swelling colouration and ovulation was not investigated, as few studies with suitable data were available (Higham et al., 2008; Setchell and Wickings, 2004; Rigaill et al., 2013). The final dataset on swelling size and female quality consisted of 7 articles across 6 species (*Macaca sylvanus*, *Mandrillus sphinx*, *Pan troglodytes*, *Papio ursinus*, *Papio anubis*, *Papio cynocephalus*) (Table A2).

### *Data for swelling size and timing of ovulation analyses*

For studies reporting summary data (*N*=11 studies), values were directly extracted from the text or figures for a) the *mean day of ovulation relative to the onset of maximal swelling size* and b) the *mean duration of the maximal swelling period*. For studies that presented data on the timing of ovulation separately for individual cycles (*N*=15 studies), we calculated mean values across all cycles based on reported values for a) the *day of ovulation relative to the onset of maximal swelling size* for each cycle (e.g. if maximal swelling size was reached 3 days prior to ovulation, the value was -3), and b) the *duration of the maximal swelling period*, i.e., the total number of days during which the swelling was maximally swollen for each cycle.

Studies were classified as using either a ‘categorical’ measure (e.g., 1=small to 3=maximal) or a ‘continuous’ measure of swelling size (e.g., area in $cm^2$ measured using calipers or estimated from digital photographs). Authors’ definitions of maximal swelling size were used throughout, or, where data were presented as graphical representations of averaged swelling size changes across the cycle (e.g. Aujard et al., 1998; Brauch et al., 2007), ‘maximal’ swelling size was taken as the point at which the peak size was reached. Authors’ estimated days of ovulation were used where available. Otherwise, ovulation was assumed to have occurred either one day after the serum LH peak, two days after the serum estradiol peak, or two days after the urinary estrone peak (Steinetz et al., 1992), unless otherwise noted. Sample sizes were taken as the number of females for which data were presented. The number of studies included in each analysis varied slightly according to available data.

### *Data for swelling size and female quality analyses*

For studies that presented information on individual female quality and swelling size, we extracted correlation coefficients (Pearson’s *r* or Spearman’s Rho). In five of the studies, correlation coefficients based on data for individual females were not reported, and requests were sent to authors, who provided either relevant *r* values (Rigaill et al., 2013; Emery and Whitten, 2003) or original data that were reanalysed using Spearman rank tests (Fitzpatrick et al., 2014; Huchard et al., 2009; Möhle et al., 2005), using average swelling size where multiple cycles were available per female. In some studies, several measures of swelling size were presented (e.g., depth, width and height), and all studies reported results for multiple measures of female quality; all values were initially extracted.

### *Phylogenetic statistical analyses*

To check for phylogenetic signal, data were collapsed to species means, and Pagel's λ was estimated (Freckleton et al., 2002; Pagel, 1999) using a dated consensus phylogeny (10kTrees version 3; Arnold et al., 2010). Maximum likelihood estimates of λ were not significantly different from 0, indicating minimal phylogenetic signal, for both the timing of ovulation and duration of maximal swelling (*Ps*>0.3). While maximum likelihood estimates of λ may be uncertain for small sample sizes (Freckleton et al., 2002), model comparison demonstrated that fixing λ=0 was strongly favoured over fixing λ=1 for both measures (AIC scores >15 units lower). Thus, the use of conventional, rather than phylogenetic, statistical methods was justified.

### *Analyses of swelling size and timing of ovulation timing*

Weighted statistical methods, which take into account differences in sample size between studies, were conducted using the R package 'weights' (version 0.75, Pasek, 2012; R Core Team, 2013). Means were weighted by the sample size (N individuals), and these weighted means were used to summarise the onset of maximal swelling in relation to ovulation and the duration of maximal swelling. Weighted T-tests were used to compare the onset and duration of maximal swelling between studies using continuous and categorical measures. Where articles presented values both for individual cycles and summarised across cycles (Higham et al., 2008; Mohle et al., 2005; Aujard et al., 1998; Nadler et al., 1985; Whitten and Russell, 1996), the summary values were used.

For four species (Barbary macaque, *Macaca sylvanus*; chimpanzee, *Pan troglodytes*; bonobo, *Pan paniscus*; olive baboon, *Papio anubis*), sufficient data on the timing of ovulation for individual cycles were available to compute the probability of ovulation for each day of the maximal swelling period (as in Barelli et al., 2007; Deschner et al., 2003). Data at the individual cycle level were available only for studies where swelling size had been recorded using categorical measures. The probability of ovulation for each day from the start of maximal swelling onwards ($T$) was calculated by dividing the number of ovulations that occurred on each day ($n_t$) by the total number of cycles being considered ($n$):

$$P\ (T = t)\ \ =\ \frac{n_t}{n}$$

where $t$ represents a specific day within the maximal swelling period.

The probability of a female being within the fertile phase of the cycle was also calculated for each day of maximal swelling (as in Barelli et al., 2007), where the fertile phase is assumed to be four days (day of ovulation and preceding three days) based on the assumption that sperm remain viable within the female reproductive tract for three days and ova remain viable for 24hrs post-ovulation (Dunson et al., 1999; Wilcox et al., 1995). Each of the days within the fertile phase was attributed a value of 1. The probability of a female being within this phase on a specific day of the cycle ($F$) was calculated as the sum of the values on a specific day of the maximal swelling period ($f_t$) divided by the total number of cycles ($n$):

$$P\ (F = f) = \frac{f_t}{n}$$

For studies that reported data for individual cycles, we calculated the percentage of ovulations falling within the maximal swelling period and the range of days of the maximal swelling period that encompassed 80% of ovulations (as in Barelli *et al.*, 2007) (Table A3).

*Analyses of swelling size and female quality*

Meta-analyses were used to estimate summary effect sizes for correlations between swelling size and measures of female quality, conducted in the R package ‘metafor’ (Viechtbauer, 2010). Random effects analyses were used, as we assumed that true effect sizes were likely to vary between studies. Effect sizes were Fisher’s Z-transformed to stabilize variance across the range of correlation coefficients (Borenstein et al., 2009). We report summary effects with 95% confidence intervals (CI) and p-values from Z-tests (Viechtbauer, 2010). Each of the seven studies reported correlations for more than one measure of female quality. Only those measures that were reported in at least four studies were analysed, i.e., i) age, ii) social rank, and iii) body condition (see Table A2 legend).

Age has been predicted to correlate positively with swelling size (Nunn et al., 2001) and was reported as a continuous measure in all studies. We predicted that social rank would correlate positively with swelling size (where highest ranking females have the highest rank value), and social rank was treated as a continuous variable. We predicted that body condition would correlate positively with swelling size, and we used body mass index (BMI) as the measure of body condition, as this measure was available for three studies (Huchard et al., 2009; Setchell and Wickings, 2004; Emery and Whitten, 2003) and could be calculated for the fourth (Möhle et al., 2005). Additional measures of body condition that were reported in only single studies, namely waist-hip ratio (Möhle et al., 2005) and mid-upper arm fat (Huchard et al., 2010), were not included.

Swelling size was reported variously as ‘height’ (i.e., vertical distance from highest to lowest point of swelling), ‘width’ (i.e., horizontal distance across widest part of swelling), ‘area’ (i.e., total swelling area as viewed from rear) or ‘depth’ (i.e., protrusion along caudal axis). We preferentially used ‘area’ as the measure of swelling size where it was reported (Huchard et al., 2009; Emery and Whitten, 2003) or could be derived from original datasets (Mohle et al., 2005), as this measure incorporates both height and width information. Where area was not available, we used ‘width’, as this measure was available for three of the four remaining studies (Fitzpatrick et al., 2014; Setchell and Wickings, 2004; Rigaill et al., 2013), while the final study reported ‘length’ (Domb and Pagel, 2001). As multiple correlation coefficients were available for two studies that did not report ‘area’ (Setchell and Wickings, 2004; Rigaill et al., 2013), we ran additional versions of the analyses using these correlations, and these versions are reported where the findings differed from those of the main analyses.

Two sources of variation between studies can impact upon observed effect sizes. First, random error can cause variation in estimates (i.e. studies might estimate the relationship between swelling size and individual quality imprecisely). Second, the ‘true’ effect to be estimated can itself vary between studies (i.e. the relationship between swelling size and individual quality actually is different when measured in different contexts). This second form of variation is termed *heterogeneity*. While random error is ubiquitous, heterogeneity must be inferred statistically. We use the $I^2$ statistic to estimate the proportion of variation in observed effect sizes that was attributable to ‘true’ variation in effects, as opposed to random error (Borenstein et al., 2009).

## Results

### *Onset of maximal swelling size and duration of maximal swelling period*

Across all studies, maximal swelling size was reached, on average, 2.75 days (SD=4.19, $N$=25) before ovulation occurred, and maximal swelling size lasted, on average, for 4.96 days (SD=3.89, $N$=24). Articles that used continuous scales to measure swelling size reported that maximal swelling size began significantly closer to ovulation (0.20 days before ovulation, SD=1.53, $N$=6) than articles using categorical scales (3.44 days before ovulation, SD=4.56, $N$=19; $T_{23.35}$=3.03, $P$<0.01), and that maximal swelling had a significantly shorter duration (1.63 days, SD=1.04, $N$=6) than articles using categorical scales (5.87 days, SD=4.00; $T_{22.83}$=4.72, $P$<0.01, $N$=18).

### *Probability of ovulation and fertility from the start of maximal swelling*

Fig. 1 provides a graphical representation of the probability of ovulation and fertility on each day during the period from the start of maximal swelling onwards for four species, where day 1 represents the first day of maximal swelling and the grey shaded area represents the mean duration of maximal swelling size.

*Percentage of ovulations that fall within the maximal swelling period*

The percentage of ovulations falling within the maximal swelling period, calculated per species, ranged from 68.4% to 93.8%, with a mean value of 81.3% (SD=10.7) (Table A3). The total number of days of the maximal swelling period that encompassed 80% of ovulations ranged from 4 to 18 days for individual species, with a mean value of 9.5 (SD=4.6).

*Female quality and swelling size*

The summary effect size for correlations between age and swelling size was not significantly different from zero (-0.06, [95% CI -0.39, 0.27], *P*=0.73, Fig. 2a). Similarly, the summary effect size for correlations with social rank was effectively zero (-0.01, [95% CI -0.20, 0.18],*P*=0.90, Fig. 2b). In contrast, the summary effect size for correlations with body condition was positive and differed significantly from zero (0.28, [95% CI 0.01, 0.56], *P*=0.04, Fig. 2c). This significant result for ‘body condition’ was replicated when using the effect size for swelling ‘depth’, rather than ‘width’, from the single study that did not report ‘area’ (0.30, [95% CI 0.02, 0.57], *P*=0.03), although not when using the ‘length’ effect size (0.22, [95% CI -0.05, 0.49], *P*=0.12).

Heterogeneity was estimated to be negligible for the effects of body condition ($I^2$ = 0%, [95% CI 0%, 88%]) and social rank ($I^2$ = 2%, [95% CI 0%, 80%]). In contrast, for age, approximately two thirds of the variance in observed effect sizes was attributable to variation in the ‘true’ strength of the relationship between age and swelling size between studies ($I^2$ = 67%, [95% CI 21%, 92%]).

## Discussion

The results confirm that exaggerated swellings are relatively reliable indicators of female fertility across the ovarian cycle. When the data from all studies were combined, maximal swelling size was reached, on average, around 3 days prior to ovulation, and the mean duration of the maximal swelling period was approximately 5 days. Given that the fertile phase is likely to last around 4 days and having shown that around 70-90% of ovulations occurred during the maximal swelling period, we assume that males that focus mating effort on maximally swollen females will have higher reproductive success than males that do not allocate mating effort according to swelling size. The available data (summarised in Fig. 1) also show that the start of the maximal swelling period tends to be accompanied by a rise in the probability of fertility. Therefore, lower ranking males, that tend to mate earlier in the swelling period than higher ranking males (Gesquiere et al. 2007; Emery Thompson, 2005), are still likely to experience some chance of paternity. Overall, these results are consistent with the hypothesis that exaggerated swellings are graded signals of fertility that function to attract multiple mating partners, while also biasing paternity towards preferred males (Nunn, 1999). Signals only need to be reliable on average to be evolutionarily stable (Johnstone and Grafen, 1993).

Our analyses do not exclude the possibility that the reliability of swelling size as a signal of fertility could be particularly low in some species (e.g., bonobos), and the analyses did not examine whether swellings that occur outside of typical ovarian cycles, such those exhibited by pregnant or lactating females (Gordon et al., 1991; Zinner and Deschner, 2000), are deceptive. However, the results did show that the type of scale used to estimate swelling size influenced the tightness of the relationship between swelling size and timing of ovulation. Studies that used categorical measures reported a significantly longer period between the onset of maximal swelling and ovulation, and a longer duration of maximal swelling, than studies that used continuous measures. For example, a study of chimpanzees, which used a categorical measure, reported that maximal swelling size was reached 7 days prior to ovulation and lasted 10 days

(Deschner et al., 2003), while another study on the same population that used a continuous measure reported that maximal swelling size coincided with the day of ovulation (Deschner et al., 2004). Categorical measures can obscure fine-scaled changes in size, potentially leading to an overestimation of the unreliability of swellings as indicators of ovarian function.

The lack of phylogenetic signal in the onset and duration of the maximal swelling period may seem surprising, given that the distribution of exaggerated swellings across the Old World primates is strongly related to phylogenetic groupings (Pagel and Meade, 2006; Nunn, 1999; Sillén-Tullberg and Møller, 1993). However, traits measured with a high degree of error are expected to exhibit low phylogenetic signal (Freckleton et al., 2002), because variation between studies can overwhelm that potentially attributable to phylogenetic relatedness. This interpretation is supported by the strong effect of measurement methodology on the recorded onset and duration of the maximal swelling period. We also expect measurement error between studies due to other methodological factors, such as differences in methods of estimating the timing of ovulation. An alternative possibility is that the onset and duration of swellings is evolutionarily more labile than the presence of swellings, allowing for swellings to function quite differently in closely related species; for example, differences in swelling size fluctuations across cycles between bonobos and chimpanzees (Zinner et al., 2004) could relate to differences in the social and sexual behaviour of these species (Ryu et al., 2015). However, because all existing bonobo studies have used categorical scales to measure swelling size, while some chimpanzee studies have used continuous scales, caution is required before drawing strong biological conclusions from the apparent divergence in swelling size fluctuations between these species.

Assuming that swelling size is a relatively reliable indicator of fertility within ovarian cycles in most species, individual males are faced with the decision of attempting to mate with a partially swollen female or waiting until swelling size increases further. The finding that low-ranking males tend to mate earlier in the cycle than high-ranking males (Gesquiere et al., 2007; Emery Thompson, 2005) does not necessarily indicate that low ranking males are being 'deceived', as individual male strategies are likely to reflect a balance between the benefits and costs of matings. Males are predicted to be choosy in their mating decisions, for example as a result of the energetic costs of courtship and sperm production (Bercovitch, 1987; Edward and Chapman, 2011) and the costs imposed by the mating tactics of other males (Fitzpartick et al., 2015). Theoretical modelling has shown that males might be deterred from responding to a strong signal by the costs of competing with same-sex conspecifics (Fawcett and Johnstone, 2003). Rather than suggesting that exaggerated swellings are 'deceptive' signals, our results instead support the idea that exaggerated swellings function to 'exploit' potential mating partners, as males are required to weigh up the costs and benefits of responding based on a probabilistic relationship between swelling size and fertility (Nunn, 1999). Conflicts of interest between the sexes have potentially led to selection for exaggerated swellings in species where female control over matings is otherwise compromised by male mating strategies (Nunn, 1999; van Schaik et al., 2000) and where males selectively allocate mating effort (Pagel, 1994).

Our analyses provide some support for the hypothesis that maximal swelling size reflects aspects of female quality other than current fertility (Pagel, 1994), as the summary effect size for correlations between body condition and swelling size was significant, even with a small number of studies. Caution is required, however, given the sensitivity of this effect size to the inclusion of specific data. However, assuming that future datasets support an association between swelling size and body condition, our analyses together suggest that males gain direct fitness benefits from allocating mating effort according to swelling size in terms of mating with females that are likely to be both fertile and in good physical condition, supporting predictions of both the graded-signal (Nunn, 1999) and reliable-indicator (Pagel, 1994) hypotheses. Whether such females also have high genetic quality cannot be determined from the current analyses, particularly as the relationship between swelling size and female reproductive success could not be investigated, due to the lack of comparable datasets. Further datasets on the relationships between female fertility, quality and swelling characteristics are required. However, as theoretical models have shown that male mate choice based on reliable signals of temporal fertility alone can maintain selection for exaggerated female ornaments (Huchard et al., 2009; Chenoweth et al., 2006), questions about the selection pressures that have favoured the evolution of exaggerated swellings can continue to be asked.

By summarising the current evidence on the relationship between exaggerated swellings and ovulation and female quality, we identify gaps in our existing knowledge and revealed heterogeneity in the available data. For example, we currently do not have detailed, quantitative information on the relationship between ovulation or female quality in less well studied species with exaggerated swellings, such as red and olive colobus monkeys (e.g., *Piliocolobus badius*, *Procolobus verus*), which appear to have evolved exaggerated swellings independently of the species included in the present study (Nunn, 1999). Furthermore, only a small number of studies have investigated the relationship between swelling size and female quality, and future empirical studies could further investigate the independent contributions of body size, body condition and swelling size to female fitness. While body mass can potentially confound relationships between body condition and swelling size (Zinner et al., 2002), in three of the four studies that reported correlation coefficients for body condition and swelling size (Emery and Whitten, 2003; Huchard et al., 2009; Setchell and Wickings, 2004), body mass was measured at a separate timepoint from the estimation of swelling size, thus reducing the chance that swelling size provided a confound in the body condition measurement. Finally, in our analyses of the relationship between age and swelling size, approximately two thirds of the variance in observed effect sizes appeared not to be attributable solely to random error, which might indicate that important moderators of this effect remain to be explored.

In conclusion, we have provided a quantitative summary of existing evidence regarding the relationship between exaggerated swellings in female Old World primates and both timing of ovulation and female quality. Our findings confirm prior evidence and hypotheses for the function of swelling size as a signal of within-cycle changes in the probability of ovulation (Nunn, 1999), supported by several additional studies published since Nunn's (1999) comprehensive review. Evidence from studies using fine-scale, continuous measures of swelling size, which have mostly been published in the past decade, suggests, however, that swellings may be more accurate signals of ovulation than implied by studies that used categorical measures. This result may have broader implications for the effect of measurement method on biological signals, and how this relates to their apparent honesty. Our meta-analyses of the relationship between swelling size and body condition provide tentative support for the hypothesis that exaggerated swellings signal also signal aspects of female quality other than current fertility (Pagel, 1994), although the analyses are limited by the small number of studies currently available. Future research into the relationship between swelling size, body condition and female fitness both within and between cycles, and across a greater variety of species, would therefore be of value in developing a greater understanding of these unusual female ornaments.

## Acknowledgements

Funding was partly provided by an ERC Advanced Grant (EVOCULTURE, ref: 232823, to Kevin Laland) and by the School of Psychology & Neuroscience, University of St Andrews.

We are grateful to Melissa Emery Thompson, Courtney Fitzpatrick, Cécile Garcia, Michael Heistermann, James Higham and Elise Huchard for providing access to original datasets, to Charles Nunn for advice on phylogenetic comparative analyses, and Kevin Laland for additional supervision. We thank Fred Bercovitch, Alan Dixson, Courtney Fitzpatrick, Elise Huchard, Mark Pagel, Alexander Weiss and two anonymous reviewers for comments on the manuscript.

## References

Aidara, D., Badawi, M., Tahiri-Zagret, C., Robyn, C. (1981). Changes in concentrations of serum prolactin, FSH, oestradiol and progesterone and of the sex skin during the menstrual cycle in the mangabey monkey (*Cercocebus atys lunulatus*). *Journal* of *Reproduction and Fertility* 62, 475–481. (doi: 10.1530/jrf.0.0620475)

Alberts, S.C., Fitzpatrick, C.L. (2012). Paternal care and the evolution of exaggerated sexual swellings in primates, *Behavioral Ecology* 23, 699–706. (doi: 10.1093/beheco/ars052)

Arnold, C., Matthews, L.J., Nunn, C.L. (2010(. The 10kTrees website: a new online resource for primate phylogeny. *Evolutionary Anthropology* 19, 114–118. (doi:10.1002/Evan.20251)

Aujard, F., Heistermann, M., Thierry, B., Hodges, J.K. (1998). Functional significance of behavioral, morphological, and endocrine correlates across the ovarian cycle in semifree ranging female Tonkean macaques. *American Journal of Primatology* 46, 285–309. (doi:10.1002/(sici)1098-2345(1998)46:4<285::aid-ajp2>3.0.co;2-8)

Barelli, C., Heistermann, M., Boesch, C., Reichard, U.H. (2007). Sexual swellings in wild white-handed gibbon females (*Hylobates lar*) indicate the probability of ovulation. *Hormones and Behavior* 51, 221–230. (doi:10.1016/j.yhbeh.2006.10.008)

Bercovitch, F.B. (1987). Reproductive success in male savanna baboons. *Behavioral Ecology and Sociobiology* 21, 163–172. (doi:10.1007/BF00303206)

Bielert, C., Anderson, C.M. (1985). Baboon sexual swellings and male response: a possible operational mammalian supernormal stimulus and response interaction. *International Journal of Primatology* 6, 377–393. (doi:10.1007/BF02736384)

Bielert, C., Busse, C. (1983). Influence of ovarian hormones on the food intake and feeding of captive and wild female chacma baboons (*Papio ursinus*). *Physiology and Behavior* 30, 103–111. (doi:10.1016/0031-9384(83)90045-8)

Blakley, G.B., Beamer, T.W., Dukelow, W.R. (1981). Characteristics of the menstrual cycle in nonhuman primates. IV. Timed mating in *Macaca nemestrina*. *Laboratory Animals* 15, 351–353. (doi: 10.1258/002367781780953059)

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. (2009). *Introduction to Meta-analysis*. Wiley, UK.

Brauch, K., Pfefferle, D., Hodges, K., Möhle, U., Fischer, J., Heistermann, M. (2007). Female sexual behavior and sexual swelling size as potential cues for males to discern the female fertile phase in free-ranging Barbary macaques (*Macaca sylvanus*) of Gibraltar. *Hormones and Behavior* 52, 375–383. (doi:10.1016/j.yhbeh.2007.06.001)

Bullock, D.W., Paris, C.A., Goy, R.W. (1972). Sexual behaviour, swelling of the sex skin and plasma progesterone in the pigtail macaque. *Journal* of *Reproduction and Fertility* 31, 225–236. (doi: 10.1530/jrf.0.0310225)

Chenoweth, S.F., Doughty, P., Kokko, H. (2006). Can non-directional male mating preferences facilitate honest female ornamentation? *Ecology Letters* 9, 179–184. (doi:10.1111/j.1461-0248.2005.00867.x)

Clutton-Brock, T. (2007). Sexual selection in males and females. *Science* 318, 1882–1885. (doi: 10.1126/science.1133311)

Clutton-Brock, T. (2009). Sexual selection in females. *Animal Behaviour* 77, 3–11. (doi:10.1016/j.anbehav.2008.08.026)

Clutton-Brock, T.H., Harvey, P.H. (1976). Evolutionary rules and primate societies, in: Bateson, P.P.G., Hinde, R.A. (Eds.), *Growing Points in Ethology*. Cambridge University Press, Cambridge, pp. 195–237.

Czekala, N., Sicotte, P. (2000). Reproductive monitoring of free-ranging female mountain gorillas by urinary hormone analysis. *American Journal of Primatology* 51, 209–215. (doi: 10.1002/1098-2345(200007)51:3<209::AID-AJP6>3.0.CO;2-6)

Dahl, J.F., Nadler, R.D., Collins, D.C. (1991). Monitoring the ovarian cycles of *Pan troglodytes* and *Pan paniscus*: a comparative approach. *American Journal of Primatology* 24, 195–209. (doi: 10.1002/ajp.1350240306)

Daspre, A., Heistermann, M., Hodges, J.K., Lee, P.C., Rosetta, L. (2009). Signals of female reproductive quality and fertility in colony-living baboons (*Papio h. anubis*) in relation to ensuring paternal investment. *American Journal of Primatology* 71, 529–538. (doi:10.1002/ajp.20684)

Deschner, T., Heistermann, M., Hodges, K., Boesch, C. (2003). Timing and probability of ovulation in relation to sex skin swelling in wild West African chimpanzees, *Pan troglodytes verus*. *Animal Behaviour* 66, 551–560. (doi:10.1006/anbe.2003.2210)

Deschner, T., Heistermann, M., Hodges, K., Boesch, C. (2004). Female sexual swelling size, timing of ovulation, and male behavior in wild West African chimpanzees. *Hormones and Behavior* 46, 204–215. (doi:10.1016/j.yhbeh.2004.03.013)

Dixson, A.F. 2012. *Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Humans, second ed*. Oxford University Press, Oxford.

Domb, L.G., Pagel, M. (2001). Sexual swellings advertise female quality in wild baboons. *Nature* 410, 204–206. (doi:10.1038/35065597)

Dunson, D.B., Baird, D.D., Wilcox, A.J., Weinberg, C.R. (1999). Day-specific probabilities of clinical pregnancy based on two studies with imperfect measures of ovulation. *Human Reproduction* 14, 1835–1839. (doi:10.1093/humrep/14.7.1835)

Edward, D.A., Chapman, T. (2011). The evolution and significance of male mate choice. *Trends in Ecology and Evolution* 26, 647–654. (doi:10.1016/j.tree.2011.07.012)

Emery, M.A., Whitten, P.L. (2003). Size of sexual swellings reflects ovarian function in chimpanzees (*Pan troglodytes*). *Behavioral Ecology and Sociobiology* 54, 340–351. (doi:10.1007/s00265-003-0648-6)

Emery Thompson, M. (2005). Reproductive endocrinology of wild female chimpanzees (*Pan troglodytes schweinfurthii*): methodological considerations and the role of hormones in sex and conception. *American Journal of Primatology* 67, 137–158. (doi:10.1002/ajp.20174)

Engelhardt, A., Hodges, J.K., Niemitz, C., Heistermann, M. (2005). Female sexual behavior, but not sex skin swelling, reliably indicates the timing of the fertile phase in wild long-tailed macaques (*Macaca fascicularis*). *Hormones and Behavior* 47, 195–204. (doi:10.1016/j.yhbeh.2004.09.007)

Fawcett, T.W., Johnstone, R.A. (2003). Mate choice in the face of costly competition. *Behavioral Ecology* 14, 771–779. (doi:10.1093/beheco/arg075)

Feldblum, J.T., Wroblewski, E.E., Rudicell, R.S., Hahn, B.H., Paiva, T., Cetinkaya-Rundel, M., Pusey, A.E., Gilby, I.C. (2014). Sexually coercive male chimpanzees sire more offspring. *Current Biology* 24, 2855–2860. (doi: 10.1016/j.cub.2014.10.039)

Fitzpatrick, C.L., Altmann, J., Alberts, S.C. (2014). Sources of variance in a female fertility signal: exaggerated estrous swellings in a natural population of baboons. *Behavioral Ecology and Sociobiology* 68, 1109–1122. (doi:10.1007/s00265-014-1722-y)

Fitzpatrick, C.L., Altmann, J., Alberts, S.C. (2015). Exaggerated sexual swellings and male mate choice in primates: testing the reliable indicator hypothesis in the Amboseli baboons. *Animal Behaviour* 104, 175–185 (doi:10.1016/j.anbehav.2015.03.019)

Freckleton, R.P., Harvey, P.H., Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* 160, 712–726. (doi:10.1086/343873)

Fürtbauer, I., Schulke, O., Heistermann, M., Ostner, J. (2010). Reproductive and life history parameters of wild female *Macaca assamensis*. *International Journal of Primatology* 31, 501–517. (doi:10.1007/s10764-010-9409-3)

Gesquiere, L.R., Wango, E.O., Alberts, S.C., Altmann, J. (2007). Mechanisms of sexual selection: Sexual swellings and estrogen concentrations as fertility indicators and cues for male consort decisions in wild baboons. *Hormones and Behavior* 51, 114–125. (doi:10.1016/j.yhbeh.2006.08.010)

Girolami, L., Bielert, C. (1987). Female perineal swelling and its effects on male sexual arousal: an apparent sexual releaser in the chacma baboon (*Papio ursinus*). *International Journal of Primatology* 8, 651–661. (doi:10.1007/bf02735782)

Gordon, T.P., Gust, D.A., Busse, C.D. (1991). Hormones and sexual behavior associated with postconception perineal swelling in the sooty mangabey (*Cercocebus torquatus atys*). *International Journal of Primatology* 12, 585–597. (doi:10.1007/BF02547671)

Graham, C.E., Preedy, J.R.K., Robinson, H., Collins, D.C. (1972). Urinary levels of estrogens and pregnanediol and plasma levels of progesterone during the menstrual cycle of the chimpanzee: relationship to the sexual swelling. *Endocrinology* 91, 13–24. (doi: http://dx.doi.org/10.1210/endo-91-1-13)

Graham, C.E., Warner, H., Misener, J., Collins, D.C., Preedy, J.R.K. (1977). The association between basal body temperature, sexual swelling and urinary gonadal hormone levels in the menstrual cycle of the chimpanzee. *Journal of Reproduction and Fertility* 50, 23–28. (doi: 10.1530/jrf.0.0500023)

Gust, D.A. (1994). Alpha-male sooty mangabeys differentiate between females' fertile and their post-conception maximal swellings. *International Journal of Primatology* 15, 289–301. (doi: 10.1007/BF02735278)

Heistermann, M., Möhle, U., Vervaecke, H., van Elsacker, L., Hodges, J.K. (1996). Application of urinary and fecal steroid measurements for monitoring ovarian function and pregnancy in the bonobo (*Pan paniscus*) and evaluation of perineal swelling patterns in relation to endocrine events. *Biology of Reproduction* 55, 844–53. (doi: 10.1095/biolreprod55.4.844)

Higham, J.P., Heistermann, M., Saggau, C., Agil, M., Perwitasari-Farajallah, D., Engelhardt, A. (2012). Sexual signalling in female crested macaques and the evolution of primate fertility signals. *BMC Evolutionary Biology* 12, 89. (doi:10.1186/1471-2148-12-89)

Higham, J.P., MacLarnon, A.M., Ross, C., Heistermann, M., Semple, S. (2008). Baboon sexual swellings: information content of size and color. *Hormones and Behavior* 53, 452–462. (doi:10.1016/j.yhbeh.2007.11.019)

Howland, B.E., Fairman, C., Butler, T.M. (1971). Serum levels of FSH and LH during the menstrual cycle of the chimpanzee. *Biology of Reproduction* 4, 101–105.

Huchard, E., Cowlishaw, G. (2011). Female-female aggression around mating: an extra cost of sociality in a multimale primate society. *Behavioral Ecology* 22, 1003–1011. (doi:10.1093/beheco/arr083)

Huchard, E., Courtiol, A., Benavides, J.A., Knapp, L.A., Raymond, M., Cowlishaw, G. (2009). Can fertility signals lead to quality signals? Insights from the evolution of primate sexual swellings. *Proceedings of the Royal Society of London B: Biological Sciences* 276, 1889–1897. (doi:10.1098/rspb.2008.1923)

Huchard, E., Raymond, M., Benavides, J., Marshall, H., Knapp, L.A., Cowlishaw, G. (2010). A female signal reflects MHC genotype in a social primate. *BMC Evolutionary Biology* 10, 96. (doi:10.1186/1471-2148-10-96)

Johnstone, J.A., Grafen, A. (1993). Dishonesty and the handicap principle. *Animal Behaviour* 46, 759–764. (doi:10.1006/anbe.1993.1253)

Johnstone, R.A., Rands, S.A., Evans, M.R. (2009). Sexual selection and condition-dependence. *Journal of Evolutionary Biology* 22, 2387–2394. (doi:10.1111/j.1420-9101.2009.01822.x)

Jurke, M.H., Hagey, L.R., Jurke, S., Czekala, N.M. (2000). Monitoring hormones in urine and feces of captive bonobos (*Pan paniscus*). *Primates* 41, 311–319. (doi: 10.1007/BF02557600)

Matsumodo-Oda, A. (1998). Injuries to the sexual skin of female chimpanzees at Mahale and their effect on behaviour. *Folia Primatologica* 69, 400–404. (doi:10.1159/000021661)

McArthur, J.W., Beitins, I.Z., Gorman, A., Collins, D.C., Preedy, J.R.K., Collins, C.E. (1981). The interrelationship between sex skin swelling and the urinary excretion of LH, estrone, and pregnanediol by the cycling female chimpanzee. *American Journal of Primatology* 1, 265–270. (doi: 10.1002/ajp.1350010303)

Möhle, U., Heistermann, M., Dittami, J., Reinberg, V., Hodges, J.K., Wallner, B. (2005). Patterns of anogenital swelling size and their endocrine correlates during ovulatory cycles and early pregnancy in free-ranging Barbary macaques (*Macaca sylvanus*) of Gibraltar. *American Journal of Primatology* 66, 351–368. (doi:10.1002/Ajp.20161)

Nadler, R.D., Dahl, J.F., Collins, D.C. (1993). Serum and urinary concentrations of sex hormones and genital swelling during the menstrual cycle of the gibbon. *Journal of Endocrinology* 136, 447–455. (doi: 10.1677/joe.0.1360447)

Nadler, R.D., Graham, C.E., Collins, D.C., Gould, K.G. (1979). Plasma gonadotropins, prolactin, gonadal steroids, and genital swelling during the menstrual cycle of lowland gorillas. *Endocrinology* 105, 290–296. (doi: http://dx.doi.org/10.1210/endo-105-1-290)

Nadler, R.D., Graham, C.E., Gosselin, R.E., Collins, D.C. (1985). Serum levels of gonadotropins and gonadal steroids, including testosterone, during the menstrual cycle of the chimpanzee (*Pan troglodytes*). *American Journal of Primatology* 9, 273–284. (doi:10.1002/ajp.1350090404)

Nunn, C.L. (1999). The evolution of exaggerated sexual swellings in primates and the graded-signal hypothesis. *Animal Behaviour* 58, 229–246. (doi:10.1006/anbe.1999.1159)

Nunn, C.L., van Schaik, C.P., Zinner, D. (2001). Do exaggerated sexual swellings function in female mating competition in primates? A comparative test of the reliable indicator hypothesis. *Behavioral Ecology* 12: 646–654. (doi: 10.1093/beheco/12.5.646)

Nyakudya, T.T., Fuller, A., Meyer, L.C.R., Maloney, S.K., Mitchell, D. (2012). Body temperature and physical activity correlates of the menstrual cycle in chacma baboons (*Papio hamadryas ursinus*). *American Journal of Physical Anthropology* 74, 1143–1153. (doi: 10.1002/ajp.22073)

Pagel, M. (1994). The evolution of conspicuous oestrus advertisement in Old World monkeys. *Animal Behaviour* 47, 1333–1341. (doi:10.1006/anbe.1994.1181)

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* 401, 887–884. (doi:10.1038/44766)

Pagel, M., Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *American Naturalist* 167, 808–825. (doi:10.1086/503444)

Pasek, J. (2012). *weights: Weighting and Weighted Statistics (R package), version 0.75*, http://CRAN.R-project.org/package=weights.

Phillips, R.S., Wheaton, C.J. (2008). Urinary steroid hormone analysis of ovarian cycles and pregnancy in mandrills (*Mandrillus sphinx*) indicate that menses, copulatory behavior, sexual swellings and reproductive condition are associated with changing estrone conjugates (E1C) and pregnanediol-3-glucuronide (PdG). *Zoo Biology* 27, 320–330. (doi: 10.1002/zoo.20192)

R Core Team. (2013). *R: a language and environment for statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Reichert, K.E., Heistermann, M., Hodges, J.K., Boesch, C, Hohmann, G. (2002). What females tell males about their reproductive status: are morphological and behavioural cues reliable signals of ovulation in bonobos (*Pan paniscus*)? *Ethology* 108, 583–600. (doi:10.1046/j.1439-0310.2002.00798.x)

Rigaill, L., Higham, J.P., Lee, P.C., Blin, A., Garcia, C. (2013). Multimodal sexual signalling and mating behavior in olive baboons (*Papio anubis*). *American Journal of Primatology* 75, 774–787. (doi:10.1002/ajp.22154)

Ryu, H., Hill, D. A. and Furuichi, T. (2015). Prolonged maximal sexual swelling in wild bonobos facilitates affiliative interactions between females. *Behaviour* 152, 285–311. (doi: 10.1163/1568539X-00003212)

Searcy, W.A., Nowicki, S. (2005). *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton University Press, Princeton.

Setchell, J.M., Charpentier, M.J.E., Bedjabaga, I.-B., Reed, P., Wickings, E.J., Knapp, L.A. (2006). Secondary sexual characters and female quality in primates. *Behavioral Ecology and Sociobiology* 61, 305–315. (doi: 10.1007/s00265-006-0260-7)

Setchell, J.M., Wickings, E.J. (2004). Sexual swelling in mandrills (*Mandrillus sphinx*): a test of the reliable indicator hypothesis. *Behavioral Ecology* 15, 438–445. (doi:10.1093/beheco/arh027)

Shaikh, A.A., Shaikh, S.A., Celaya, C.L., Gomez, I. (1982). Temporal relationship of hormonal peaks to ovulation and sex skin deturgescence in the baboon. *Primates* 23, 444–452. (doi: 10.1007/BF02381326)

Sillén-Tullberg, B., Møller, AP. (1993). The relationship between concealed ovulation and mating systems in anthropoid primates: a phylogenetic analysis. *American Naturalist* 141, 1–25. (doi:10.1086/285458)

Steinetz, B.G., Ducrot, C., Randolph, C., Mahoney, C.J. (1992). Determination of the time of ovulation in chimpanzees by measurement of LH, estrone sulphate, and pregnanediol s-glucuronide in urine: comparison with serum hormone patterns. *Journal of Medical Primatology* 21, 239–245.

Thierry, B., Heistermann, M., Aujard, F., Hodges, J.K. (1996). Long-term data on basic reproductive parameters and evaluation of endocrine, morphological, and behavioral measures for monitoring reproductive status in a group of semifree-ranging Tonkean macaques (*Macaca tonkeana*). *American Journal of Primatology* 39, 47–62. (doi: 10.1002/(SICI)1098-2345(1996)39:1<47::AID-AJP4>3.0.CO;2-S)

Thomson, J.A., Hess, D.L., Dahl, K.D., Iliff-Sizemore, S.A., Stouffer, R.L., Wolf, D.P. (1992). The Sulawesi crested black macaque (*Macaca nigra*) menstrual cycle: changes in perineal tumescence and serum estradiol, progesterone, follicle-stimulating hormone, and luteinizing hormone levels. *Biology of Reproduction* 46, 879–884. (doi: 10.1095/biolreprod46.5.879)

van Schaik, C.P., Hodges, J.K., Nunn, C.L. (2000). Paternity confusion and the ovarian cycles of female primates, in: van Schaik, C.P., Janson, C.H. (Eds.), *Infanticide by Males and its Implications*. Cambridge University Press, Cambridge, pp. 361–387.

van Schaik, C.P., van Noordwijk, M.A., Nunn, C.L. (1999). Sex and social evolution in primate, in: Lee, P.C. (Ed.), *Comparative Primate Socioecology*. Cambridge University Press, Cambridge, pp. 204–240.

Viechtbauer, W. (2010.) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36, 1–48.

Whitten, P.L., Russell, E. (1996). Information content of sexual swellings and fecal steroids in sooty mangabeys (*Cercocebus torquatus atys*). *American Journal of Primatology* 40, 67–82. (doi: 10.1002/(SICI)1098-2345(1996)40:1<67::AID-AJP5>3.0.CO;2-Z)

Wilcox, A.J., Weinberg, C.R., Baird, D.D. (1995). Timing of sexual intercourse in relation to ovulation: effects on the probability of conception, survival of the pregnancy, and sex of the baby. *New England Journal of Medicine* 333, 1517–1521. (doi:10.1056/NEJM199512073332301)

Wildt, D.E., Doyle, U., Stone, S.C., Harrison, R.M. (1977). Correlation of perineal swelling with serum ovarian hormone levels, vaginal cytology and ovarian follicular development during the baboon reproductive cycle. *Primates* 18, 261–270. (doi: 10.1007/BF02383104)

Young, C., Majolo, B., Heistermann, M., Schülke, O., Ostner, J. (2013). Male mating behaviour in relation to female sexual swellings, socio-sexual behaviour and hormonal changes in wild Barbary macaques. *Hormones and Behavior* 63, 32–39. (doi:10.1016/j.yhbeh.2012.11.004)

Zinner, D.P., Deschner, T. (2000). Sexual swellings in female hamadryas baboons after male take-overs: “deceptive” swellings as a possible female counter-strategy against infanticide. *American Journal of Primatology* 52, 157–168. (doi:10.1002/1098-2345(200012)52:4<157::AID-AJP1>3.0.CO;2-L)

Zinner, D.P., Alberts, S.C., Nunn, C.L., Altmann, J. (2002). Significance of primate sexual swellings. *Nature* 420, 142–143. (doi: 10.1038/420143a)

Zinner, D.P., Nunn, C.L., van Schaik, C.P., Kappeler, P.M. (2004). Sexual selection and exaggerated sexual swellings of female primates, in: Kappeler, P.M., van Schaik, C.P. (Eds.), *Sexual Selection in Primates*. Cambridge University Press, Cambridge, pp. 71–89.

## Figure legends

**Figure 1** Probability of ovulation (solid line) and fertility (dashed line) across the ovarian cycle relative to the onset of maximal swelling size (day 1 is first day of maximal swelling; grey shading represents mean duration of maximal swelling) for Barbary macaques (*Macaca sylvanus*; *N*=19 cycles; Möhle et al., 2005; Young et al., 2013), chimpanzees (*Pan troglodytes*; *N*=46 cycles; Deschner et al., 2003; Graham et al., 1972; Graham et al., 1977; Howland et al., 1971; McArthur et al., 1981; Nadler et al., 1985), bonobos (*Pan paniscus*; *N*=34 cycles; Heistermann et al., 1996; Jurke et al., 2000; Reichert et al., 2002), and olive baboons (*Papio anubis*; *N*=22 cycles; Higham et al., 2008; Daspre et al., 2009). All of the studies used to construct Fig. 1 reported categorical measures of swelling size. The mean duration of maximal swelling size was calculated from the cited studies. Cycle length data were taken from van Schaik et al., 1999.

**Figure 2** Forest plots showing effect sizes (Fisher's Z transformed correlation coefficients), sample sizes and summary effects from random-effects meta analyses of the relationship between swelling size and a) age, b) social rank, and c) body condition.

**Figure 1**

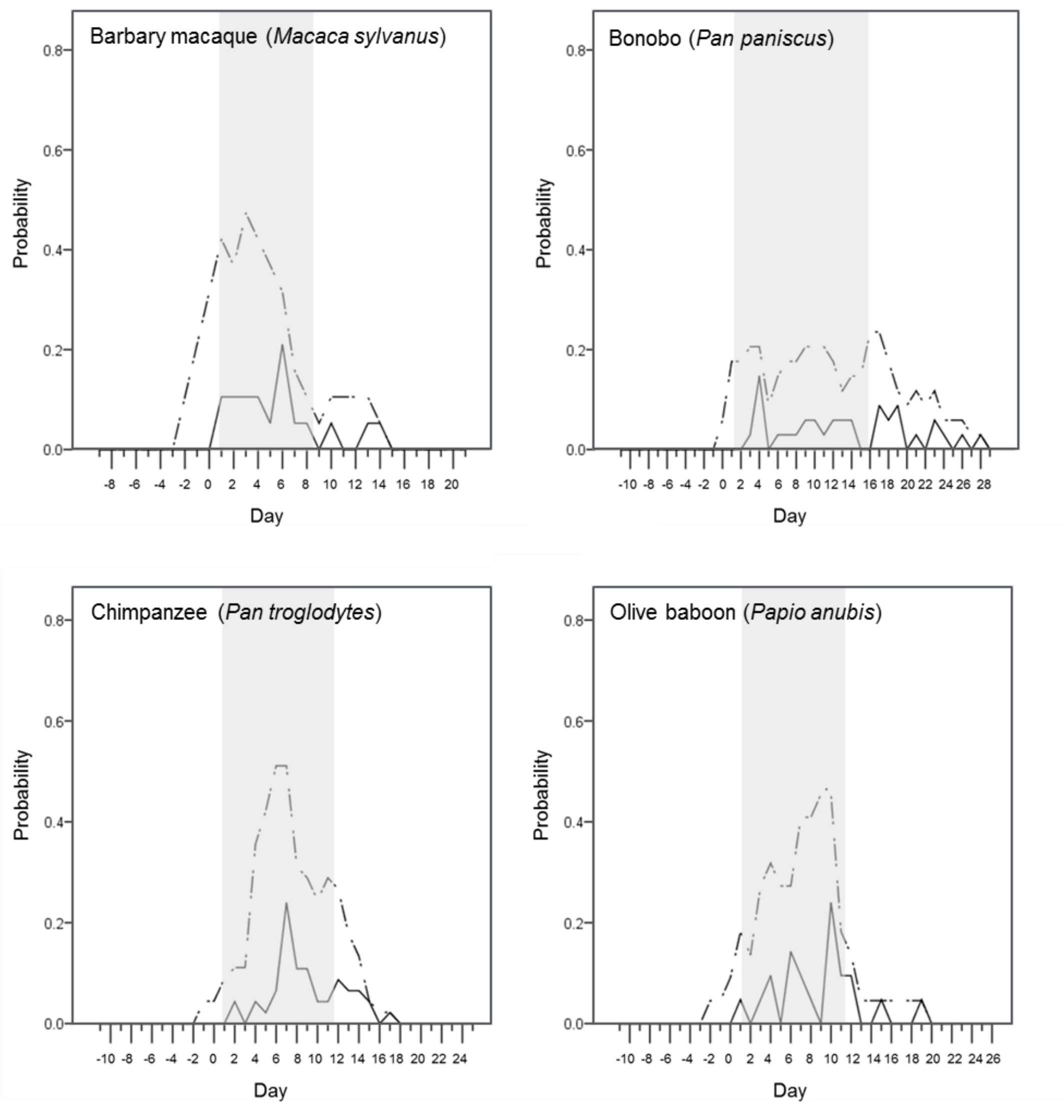
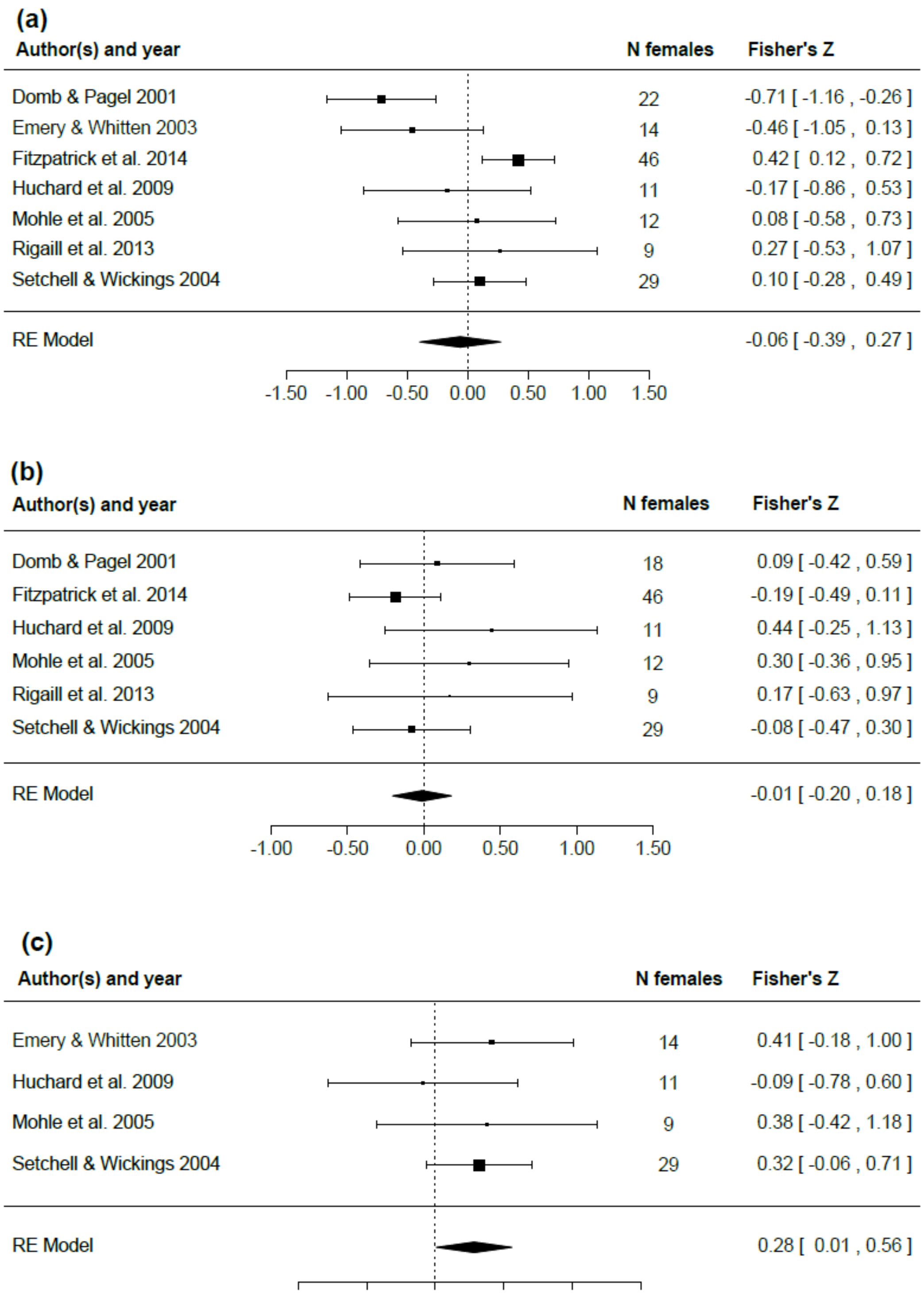

Barbary macaque (*Macaca sylvanus*)
Bonobo (*Pan paniscus*)
Chimpanzee (*Pan troglodytes*)
Olive baboon (*Papio anubis*)
Probability
0.0
0.2
0.4
0.6
0.8
Day
-10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 18 20 22 24 26 28

**Figure 2**


(a)
Author(s) and year
N females
Fisher's Z
Domb & Pagel 2001
22
-0.71 [ -1.16 , -0.26 ]
Emery & Whitten 2003
14
-0.46 [ -1.05 , 0.13 ]
Fitzpatrick et al. 2014
46
0.42 [ 0.12 , 0.72 ]
Huchard et al. 2009
11
-0.17 [ -0.86 , 0.53 ]
Mohle et al. 2005
12
0.08 [ -0.58 , 0.73 ]
Rigaill et al. 2013
9
0.27 [ -0.53 , 1.07 ]
Setchell & Wickings 2004
29
0.10 [ -0.28 , 0.49 ]
RE Model
-0.06 [ -0.39 , 0.27 ]
-1.50
-1.00
-0.50
0.00
0.50
1.00
1.50
(b)
Author(s) and year
N females
Fisher's Z
Domb & Pagel 2001
18
0.09 [ -0.42 , 0.59 ]
Fitzpatrick et al. 2014
46
-0.19 [ -0.49 , 0.11 ]
Huchard et al. 2009
11
0.44 [ -0.25 , 1.13 ]
Mohle et al. 2005
12
0.30 [ -0.36 , 0.95 ]
Rigaill et al. 2013
9
0.17 [ -0.63 , 0.97 ]
Setchell & Wickings 2004
29
-0.08 [ -0.47 , 0.30 ]
RE Model
-0.01 [ -0.20 , 0.18 ]
-1.00
-0.50
0.00
0.50
1.00
1.50
(c)
Author(s) and year
N females
Fisher's Z
Emery & Whitten 2003
14
0.41 [ -0.18 , 1.00 ]
Huchard et al. 2009
11
-0.09 [ -0.78 , 0.60 ]
Mohle et al. 2005
9
0.38 [ -0.42 , 1.18 ]
Setchell & Wickings 2004
29
0.32 [ -0.06 , 0.71 ]
RE Model
0.28 [ 0.01 , 0.56 ]

# Appendix

**Table A1** Summary of studies containing data on swelling size and ovulation

| Species | Sample size | Day of ovulation | Duration of maximal swelling | Swelling scale | Sources |
|---|---|---|---|---|---|
| Sooty mangabey (*Cercocebus atys*) | 6 | -1.0 | 1 | b | Aidara et al., 1981 |
| | 4 | -2.0 | NA | a | Gust, 1994 |
| | 3 | -4.0 | 5.0 | a | Whitten and Russell, 1996*[1] |
| Pig-tailed macaque (*Macaca nemestrina*) | NA | -5.5 | 5.0 | a | Blakley et al., 1981 |
| Barbary macaque (*Macaca sylvanus*) | 11 | 0.0 | 1.0 | b | Brauch et al., 2007 |
| | 5 | 3.0 | 1.0 | b | Möhle et al., 2005*[2] |
| | 10 | -5.2 (4.6) | 6.1 (3.4) | a | Young et al., 2013* |
| Celebes black macaque (*Macaca nigra*) | 19 | 0.0 | 3.0 | b | Higham et al., 2012[3] |
| Tonkean macaque (*Macaca tonkeana*) | 6 | -3.0 | 3.0 | a | Aujard et al., 1998*[4] |
| | 4 | -0.1 (1.9) | 2.1 (0.4) | a | Thierry et al., 1996*[5] |
| Mandrill (*Mandrillus sphinx*) | 3 | 0.0 | 2.0 | a | Phillips and Weaton, 2008[6] |
| Bonobo (*Pan paniscus*) | 5 | -10.2 (7.4) | 12.8 (8.0) | a | Heistermann et al., 1996* |
| | 1 | -3.0 (0.0) | 9.5 (3.5) | a | Jurke et al., 2000* |
| | 8 | -14.3 (6.5) | 15.0 (6.7) | a | Reichert et al., 2002* |
| Chimpanzee (*Pan troglodytes*) | 12 | -7.4 (2.8) | 9.9 (3.3) | a | Deschner et al., 2003*[7] |
| | 8 | -4.0 | 4.0 | a | Emery and Whitten, 2003 |
| | 45 | 1.0 | 4.0 | a | Emery Thompson, 2005[6] |
| | 3 | -9.3 (5.7) | 9.7 (7.8) | a | Graham et al., 1972* |
| | 2 | -10.0 (2.0) | 8.3 (3.1) | a | Graham et al., 1977* |
| | 4 | -7.5 (6.4) | 11.8 (7.9) | a | Howland et al., 1971* |
| | 1 | -12.0 | 16.0 | a | McArthur et al., 1981* |
| | 7 | -11.0 | 14.5 (2.0) | a | Nadler et al., 1985*[8] |
| Olive baboon (*Papio anubis*) | 9 | -10.0 (4.1) | 10 (4.6) | a | Daspre et al., 2009* |
| | 7 | -3.0 | 1.0 | b | Higham et al., 2008[9] |
| Yellow baboon (*Papio cynocephalus*) | 88 | -2.0 | 4.0 | a | Gesquire et al., 2007[6,10] |

Sample size = number of individuals. Day of ovulation (defined as day zero) is reported relative to onset of maximal swelling and is the mean number of days, with SDs in brackets where available. Duration of maximal swelling period is the mean number of days, with SDs in brackets where available. Swelling scale: a = categorical, b= continuous. Studies marked with an asterisk presented data on the timing of ovulation for individual cycles.

[1] This study presented data both summarised across cycles (values given in the table) and at the cycle level (ovulation: mean = -2.4, SD = 4.1; duration: mean =7.0, SD = 1.2; n = 4).

[2] Ovulation was assumed to have occurred three days before the fecal progesterone rise (based on Barelli et al., 2007). This study presented data both summarised across cycles (values in table) and at cycle level (ovulation: mean = -2.6, SD = 2.2; duration: mean = 7.1, SD = 2.1; n = 6).

[3] Ovulation was taken as the first day of the 'ovulation window' defined by the authors.

[4] This study presented data both summarised across cycles (values given in table) and at cycle level (ovulation: mean = -5.2, SD = 1.7; duration: mean = 8.6, SD = 2.0; n = 6).

[5] Ovulation assumed to have occurred one day before rise in urinary progesterone levels (based on Deschner et al., 2003).

[6] Data on hormones levels and swelling size are aligned to detumescence or onset of menses, rather than day of ovulation.

[7] Data using continuous scale are also available for the same subjects (ovulation: mean = 0.0, SD = NA; duration: mean = 1.0, SD = NA; n=12; Deschner et al., 2004).

[8] This study presented data both summarised across cycles (values given in table) and at cycle level (ovulation: mean = -13.5, SD = 0.7; duration: mean = 15.5, SD = 0.7; n = 2).

[9] Data for individual cycles were provided by the author (ovulation: mean = -6.2, SD = 3.4; duration: mean = 9.3, SD = 3.8; n=7).

[10] Ovulation was assumed to have occurred two days after the fecal estrogen peak (based on Emery Thomson, 2005).

**Table A2** Summary of studies containing data on swelling size and measures of individual female quality.

| Measure | Species | Sources |
|---|---|---|
| Age | Barbary macaque (*Macaca sylvanus*) | Möhle et al., 2005 |
| | Mandrill (*Mandrillus sphinx*) | Setchell and Wickings, 2004 |
| | Chimpanzee (*Pan troglodytes*) | Emery and Whitten, 2003 |
| | Chacma baboon (*Papio ursinus*) | Huchard et al., 2009 |
| | Olive baboon (*Papio anubis*) | Domb and Pagel, 2001; Rigaill et al., 2013 |
| | Yellow baboon (*Papio cynocephalus*) | Fitzpatrick et al., 2014 |
| Social rank | Barbary macaque (*Macaca sylvanus*) | Möhle et al., 2005 |
| | Mandrill (*Mandrillus sphinx*) | Setchell and Wickings, 2004 |
| | Chacma baboon (*Papio ursinus*) | Huchard et al., 2009 |
| | Olive baboon (*Papio anubis*) | Domb and Pagel, 2001; Rigaill et al., 2013 |
| | Yellow baboon (*Papio cynocephalus*) | Fitzpatrick et al., 2014 |
| Body condition | Barbary macaque (*Macaca sylvanus*) | Möhle et al., 2005 |
| | Mandrill (*Mandrillus sphinx*) | Setchell and Wickings, 2004 |
| | Chimpanzee (*Pan troglodytes*) | Emery and Whitten, 2003 |
| | Chacma baboon (*Papio ursinus*) | Huchard et al., 2009 |

An additional study was excluded (Gesquiere et al., 2007), as more recent data were available from the same population (Fitzpatrick et al., 2014; also see Fitzpatrick et al., 2015). A study that reported the relationship between swelling size and genetic diversity, immune status and levels of parasitism (Setchell et al., 2006) was excluded, as these variables were not measured in any other studies. A study that reported the relationship between swelling size and a categorical measure of number of offspring (‘parity’; i.e., nulliparous, primiparous, multiparous) was excluded (Higham et al., 2008), as an insufficient number of studies with comparable measures of parity was available. While three other studies reported results for categorical measures of parity (Emery and Whitten, 2003; Huchard et al., 2009; Möhle et al., 2005), two of these studies had small sample sizes of females in specific parity categories (≤3; Huchard et al., 2009; Möhle et al., 2005) and were therefore deemed unsuitable for inclusion. Correlation coefficients for the relationship between swelling size and parity could be extracted for two studies (Fitzpatrick et al., 2014; Setchell and Wickings, 2004); however, this number of studies was too small for a meta-analysis, and an additional coefficient for number of offspring per year (Domb and Pagel, 2001) was not considered sufficiently comparable for inclusion. The correlation coefficients used in the meta-analysis are not always directly comparable to the statistics reported in original articles, as some authors analysed the relationship between swelling size and several predictor variables simultaneously using linear mixed models or compared sets of models that contained several predictor variables (Fitzpatrick et al., 2014; Huchard et al., 2009; also see Huchard et al., 2010); for these studies, original data were available and were converted to standardised measures for the meta-analyses. One study (Rigaill et al., 2013) provided correlations for pre-fertile, fertile and post-fertile cycle phases, from which we used correlations from the fertile phases only. The data used in the meta-analyses are available in the Electronic Supplementary Material.

**Table A3** Percentage of ovulations falling within the maximal swelling period, and days of the maximal swelling period that encompass 80% of ovulations, for each species where data on the timing of ovulation for individual cycles were available.

| Species | Sample size | Percentage of ovulations in maximal swelling period | Days of maximal swelling period that encompass 80% of ovulations (N days) | Sources |
|---|---|---|---|---|
| Sooty mangabey (*Cercocebus atys*) | 7 | 71.4 | 1st - 3rd (4) | Whitten and Russell, 1996 |
| Barbary macaque (*Macaca sylvanus*) | 19 | 68.4 | 1st – 8th (8) | Möhle et al., 2005; Young et al., 2013 |
| Tonkean macaque (*Macaca tonkeana*) | 16 | 93.8 | 1st – 8th (8) | Aujard et al., 1998; Thierry et al., 1996 |
| Bonobo (*Pan paniscus*) | 34 | 76.5 | 4th – 21st (18) | Heistermann et al., 1996; Jurke et al., 2000; Reichert et al., 2002 |
| Chimpanzee (*Pan troglodytes*) | 46 | 91.3 | 6th – 14th (9) | Deschner et al., 2003; Graham et al., 1972; Graham et al., 1977; Howland et al., 1971; McArthur et al., 1981; Nadler et al., 1985 |
| Olive baboon (*Papio anubis*) | 22 | 86.4 | 6th – 15th (10) | Daspre et al., 2009; Higham et al., 2008 |

Sample size = number of cycles. First day of the maximal swelling period is defined as day 1.