Title: The Goldilocks Effect: Female geladas in mid-sized groups have higher fitness

Authors: Elizabeth Tinsley Johnson[1*†‡], Noah Snyder-Mackler[2,3*†], Amy Lu[4], Thore J. Bergman[5,6], Jacinta C. Beehner[1,5].

Affiliations:

[1]Department of Anthropology, University of Michigan.

[2]Department of Psychology, University of Washington.

[3]Center for Studies in Demography & Ecology, University of Washington.

[4]Department of Anthropology, Stony Brook University.

[5]Department of Psychology, University of Michigan.

[6]Department of Ecology and Evolutionary Biology, University of Michigan.

*Correspondence to: etinsley@msu.edu and nsmack@uw.edu.

†These authors contributed equally to this work.

‡Present address: Human Biology Program and Department of Integrative Biology, College of Natural Science, Michigan State University.

Abstract: The cost-to-benefit ratio of group-living is thought to vary with group size: individuals in “optimal” groups should have higher fitness than individuals in groups that are too large or small. However, the relationship between group size and individual fitness has been difficult to establish, a gap we address here in the gelada. We demonstrate group size effects on the production of surviving offspring and on female mortality rates, which are largely explained by group-size variation in infanticide risk and foraging competition. We also identify a mechanism by which females may alter group size: in large groups, females groomed with less than half of their group, increasing the likelihood of fissions. Our findings provide insight into how and why group size shapes fitness in long-lived species.

Keywords: optimal group size, fitness, folivore paradox, infanticide, reproduction, survival

Main Text:

Variability in group size within species reflects a delicate balance of the costs and benefits of group living. For example, in large groups the costs of group-living (e.g., foraging competition: *1*–*3*) may begin to outweigh the benefits (e.g., decreased predation risk: *1*, *4*, *5*), to the detriment of individual group members. Individuals in "optimal" groups, by contrast, should have the highest lifetime reproductive success (i.e., fitness: *6*) compared to others in the population (*7*). Although the link between group size and fitness has long been suspected, it has been challenging to actually demonstrate with empirical data. This is, in part, because fitness in long-lived species is difficult to measure. Instead, short-term measures of reproductive performance (e.g., birth rate) are often used as proxies for fitness. But, perhaps more problematic is that the accuracy of measures of fitness (which are more accurate over a longer period of time) are inversely related to the accuracy of measures of group size (which are more accurate over a shorter period of time). To circumvent this problem, we developed a novel measure of reproductive performance that accounts for changes in group size.

Here, we used a wild primate, geladas (*Theropithecus gelada*), as a test species to identify whether we have evidence for an optimal group size for females and, if so, whether we can identify the selective forces that shape it. Geladas are long-lived, folivorous primates that live in social groups (hereafter, "units") that vary in size from 1 to 13 adult females – variation that is present even within the same population (*8*). This variability may be due to ecological pressures (e.g., due to the seasonal environment: *9*) as well as social pressures. For instance, infanticide is the leading cause of infant mortality: immigrant males kill over half of the dependent infants in a unit (*10*). Moreover, larger units experience more takeovers than smaller units (*11*). Females in larger units might therefore incur disproportionately higher costs than females in smaller units.

We analyzed 9 years of data from 33 units that varied in size (Fig. 1), often month-to-month. To address female fitness, we used two independent measures: (a) *Reproductive performance*: A monthly binary score, where “successful” females (those that produced an infant that survived to 18 months of age, the mean age at weaning for this population: *12*) received 1 point for each month of pregnancy and lactation (6 months of gestation + 18 months of lactation = 24 successful months). If an infant died before reaching 1.5 years of age, no points were given to the months of gestation/lactation prior to the infant death. Importantly, this monthly reproductive performance variable is associated with a precise measure of group size; (b) *Adult female death rate*: The number of adult female deaths out of the number of individual females observed in each unit size each year.

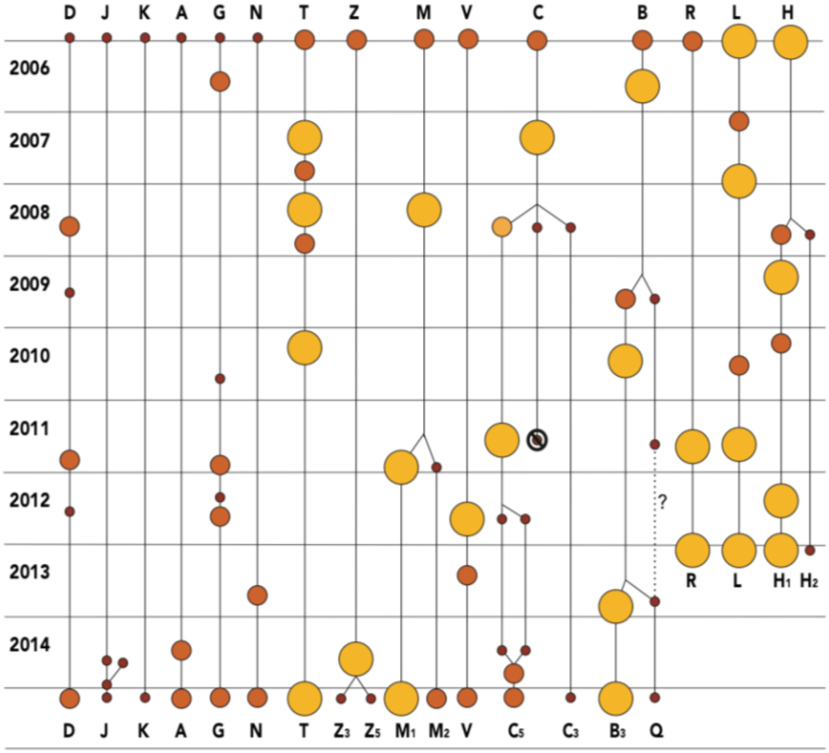

Fig. 1. Changes in unit size over time. Founding units (where observation started in 2006) are coded by letters and arranged along the top axis according to initial size category (which is also indicated by the size and color of the node). Size categories reflect the observed range in variation in unit size during the study period (small = 1-4, in dark red; mid-sized = 5-7, in orange; large = 8+, in yellow), and were used for visualization purposes only. Fissions and fusions are indicated by diagonal lines.

For our analyses, we defined unit size as the number of adult females in a unit (which is also highly correlated with the total number of individuals in a unit, Fig. S1). Although units in this population can range in size from 1-13 adult females, females in units in the middle of this range (i.e., between 5-7 adult females, hereafter “mid-sized” units) demonstrated the highest fitness. First, females in mid-sized units had the highest reproductive performance (maximum females^2 estimate = -11.22 +/- 3.30 SE, *p*-value = 6.7×10^{-4}, Fig. 2A), with 26.4% more productive months than females in small units and 17.2% more productive months than females in large units (Fig. 2B). Second, females in mid-sized units had the lowest mortality (maximum females^2 estimate = 13.52 +/- 3.01 SE, *p*-value = 7.07 x 10^{-6}, Fig. 2C), with 51.8% lower mortality than females in small units and 29.4% lower mortality than females in large units (Fig. 2D).

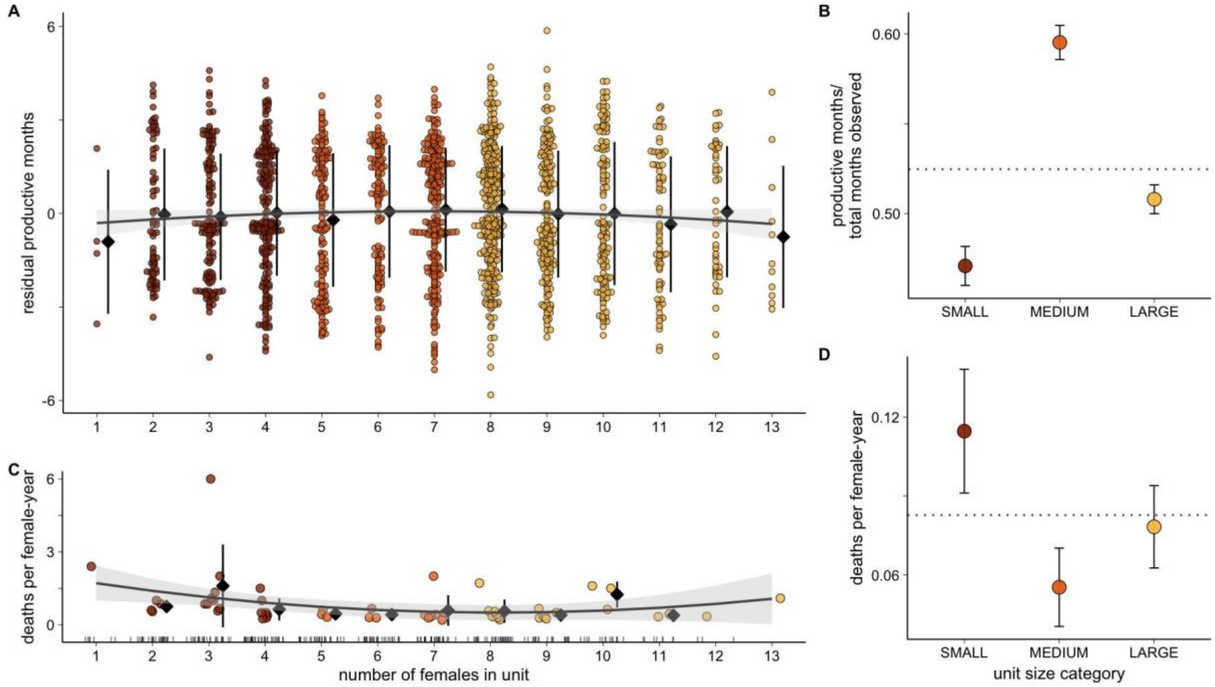

Fig. 2. Females in mid-sized units had the highest fitness. (A) Residual productive months (controlling for age) by unit size (number of adult females). Each circle indicates the summed reproductive performance for a female over a 6-month period (n = 1677 observations from n = 185 females; binning was done for visualization purposes). Black diamonds indicate the mean for each unit size, and error bars indicate the standard deviation around the mean; **(B)** Total productive months divided by total months observed for all females by unit size category. The dotted horizontal line = 0.53 (the population mean); **(C)** Female death rates (deaths per female-year observed) by unit size (*n* = 32 distinct units; each point represents a unit-size-year where at least one death occurred, i.e., for each year, we summed the number of deaths for each unit at a specific unit size and controlled for the total female-years observed at that size, *n* = 55 unit-size-years). Unit-size-years where no deaths occurred (n = 250 unit-size-years) are represented as tick marks along the x-axis. Black diamonds indicate the mean death rates for each unit size, and error bars indicate the standard deviation around the mean; **(D)** Female death rates by unit size category. The dotted horizontal line = 0.08 (the population mean). Unit size categories were used for visualization purposes only (see *Supplementary Methods*).

If mid-sized units are optimal for female fitness, we then asked why. Specifically, can social (male takeovers) and/or ecological (female feeding competition) pressures explain why females living in mid-sized units had higher fitness? Geladas - folivores that feed on widely-dispersed food resources (*13*, *14*) - should face low feeding competition and should, correspondingly, live in large groups for predator protection (*15*). Yet, this is not the pattern observed across numerous folivore taxa (the “Folivore Paradox”: *15*; i.e., *16*). The leading hypothesis for why many folivores do not live in larger groups is that smaller groups are less attractive targets for male takeovers and thus infanticide (*17*–*19*), although within-group feeding competition may also play a role (*20*). Therefore, we first examined how infanticide risk, as measured by both takeover frequency and infanticide rates, varied across unit size.

As reported previously (*10*, *21*), takeover frequency increased with unit size (maximum females estimate = 0.51 +/- 0.20 SE, *p*-value = 1.29×10^{-2}; Fig. 3A). However, infanticides, the leading cause of infant death (12.3% of all infants born during this study period died of infanticide, accounting for 50.8% of all dependent infant deaths), did not follow this same pattern. Infanticide rates actually decreased with unit size overall (maximum females estimate = -18.35 +/- 6.51 SE, *p*-value = 4.8×10^{-3}; Fig. 3B), and were lowest in mid-sized units (maximum females^2 estimate = 24.40 +/- 5.85 SE, *p*-value = 3.0×10^{-5}; Fig. 3B). Thus, while females in large units experienced more frequent takeovers and the highest number of infanticides (17 out of 33 observed infanticides occurred in large units), females in small units experienced the second highest number of infanticides (10 out of 33) such that females in mid-sized units experienced the lowest infanticide rates (rates were identical for both small and large units at 0.14 infanticides / birth; compared to 0.08 infanticides / birth for mid-sized units).

Although takeovers of small units were less frequent, when they did occur, immigrant males killed disproportionately more infants than when they took over mid-sized units. In this population, male takeovers dramatically alter female reproductive patterns (*9*, *22*); here we show that females in mid-sized units have a selective advantage over females in smaller or larger units when it comes to these male-mediated birth patterns.

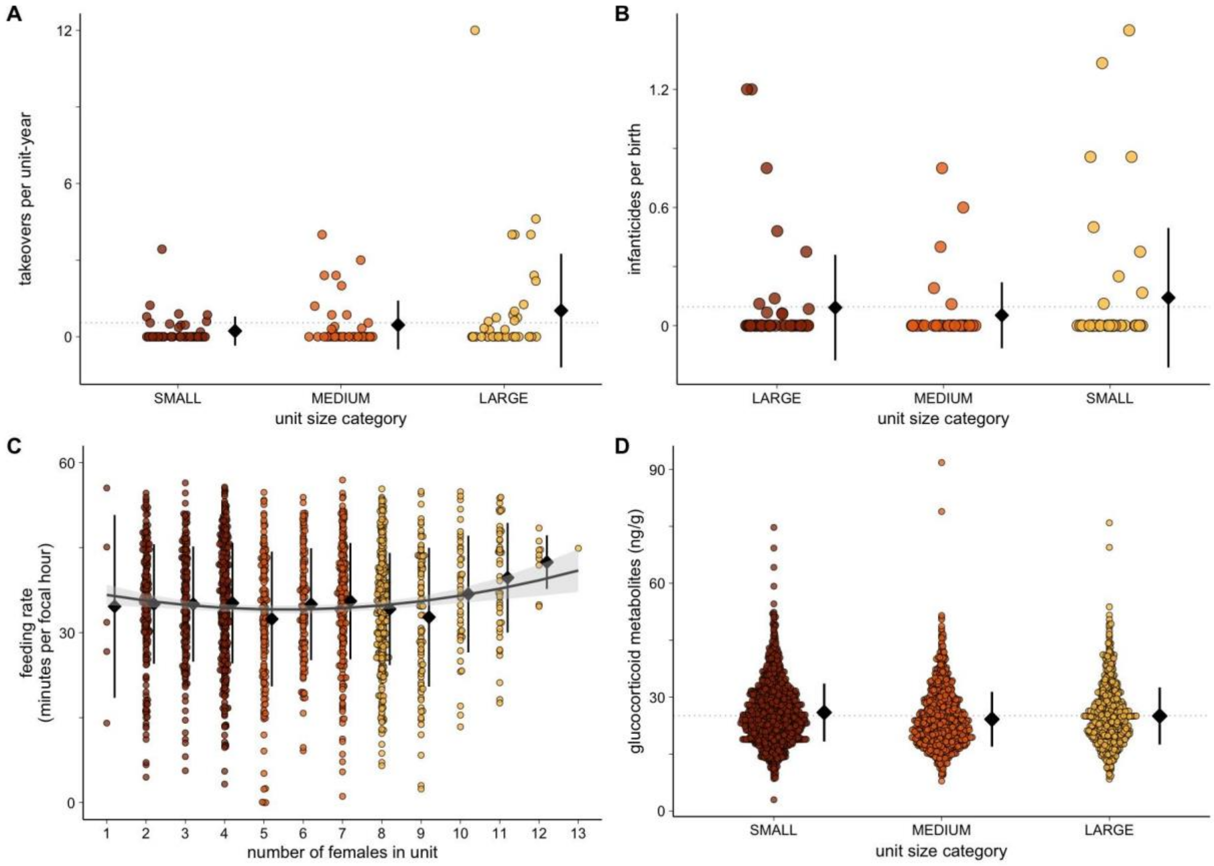

Fig. 3. Both social and ecological factors explain why mid-sized units are optimal. (A) Number of takeovers per unit-year observed. Each circle represents the takeover rate for each unit at a specific size category. The dotted horizontal line indicates the overall mean takeover rate (0.55 takeovers per unit-year); **(B)** The infanticide rates (number of infanticides per infants born for each unit-size-year) for each unit size category in a given year. Unit-size-years where no infants were born were excluded from the figure (n = 32 unit-size-years out of 97 total). The dotted horizontal line indicates the overall mean infanticide rate (0.01 infanticides per birth); **(C)** Feeding rates for each focal female for each month observed (n = 4,350 female-months). Rates were calculated by dividing the total minutes observed feeding by the focal hours observed; **(D)** Adult female glucocorticoid metabolites (ng/g) by unit size category (n = 3835 samples). The dotted horizontal line indicates the overall mean (25.11 ng/g). For all figures, black diamonds reflect the mean for each category, with bars representing the standard deviation.

Socioecological theory predicts that female competition within a group should select for smaller group sizes, while female competition between groups should select for larger group sizes (*23*). We therefore examined whether we could detect increased competition in two parameters: the amount of time females spent foraging (*24*) and female energetic condition as measured by fecal glucocorticoids (*25*). Specifically, we considered whether foraging effort (as measured by time spent feeding) or glucocorticoid levels (a class of hormones that rise in response to increased metabolic demands) varied across unit size.

Despite feeding primarily on dispersed grasses (*13*), we found evidence that geladas may experience feeding competition associated with unit size. Specifically, unit size predicted foraging effort: females in large units spent 53.89% more time foraging than females in small units (large unit category estimate = 4.01 +/- 1.92 SE, z-value = 2.09, *p*-value = 4.2 x 10^{-2}; Fig. 3C), and females in mid-sized units spent the least amount of time foraging overall (maximum females^2 estimate = 2.30 +/- 0.72 SE, z-value=3.18, *p*-value = 1.5 x 10^{-3}; Fig. 3C).

We found no such group size association with glucocorticoid metabolites (maximum females estimate = -0.91 +/- 0.61 SE, *p*-value = 0.14; maximum females^2 estimate = 0.67 +/- 0.44 SE, *p*-value = 0.13). Previous analysis in this population have indicated that temperature, and not the availability of green grass, most strongly predicts glucocorticoid levels (*9*, *26*). Therefore, thermoregulatory demands, which presumably affect all individuals equally across units of different size, may constrain metabolic needs more than food does. In addition, individual characteristics (e.g., gregariousness) are likely important predictors of variation in glucocorticoid levels, especially where more social partners may provide certain benefits (e.g., thermoregulatory benefits: *27*). Both of these hypotheses warrant further investigation.

Taken together, our results show that females in mid-sized units appear to display optimal fitness as the result of social and ecological pressures. These findings beg the question: what, if anything, can females in seemingly “suboptimal” units do to improve their situation? As the philopatric sex, female geladas have limited options for “choosing” their unit size. Nevertheless, females have been observed to change their unit size via unit transfers, fissions, and fusions. Unit transfers were rare (n=2; 0.003 transfers per female-year) and were only observed in larger units. Fusions were also relatively rare and only observed in smaller units (n=3; 0.004 fusions per female-year). In contrast, fissions were 3 times more common than fusions; females in larger units tended to fission into 2 (or 3) daughter units (9 cases; 0.01 fissions per female-year). Thus, females in suboptimal groups may achieve a more optimally-sized group by forming a smaller unit when they are in large units or by forming a larger unit when they are in small units.

We next tested how females might induce these changes — hypothesizing that individual-level behaviors may precipitate unit fissions. Individuals in larger groups are less likely to maintain social ties with everyone else in the group -- thus increasing the likelihood of a fission (*28–31*). In particular, female geladas in small and large units that spend more time feeding, may be more constrained in the amount of time they can devote to grooming (the primary means by which primates maintain social relationships, *32*), ultimately limiting the number of social ties in these units. To test this possibility, we first examined whether grooming time and/or the number of grooming partners varied by unit size.

Time spent grooming did not vary according to unit size (maximum females estimate = 1.35 +/- 4.43 SE; *p*-value = 0.76; maximum females^2 estimate = -3.47 +/- 3.74 SE; *p*-value = 0.36; Fig. 4A). This suggests that, even in the face of increasing foraging effort (Fig. 3C), gelada females in all units conserve a set amount of social time. However, while grooming time did not increase with unit size, the number of grooming partners did (range: 0-6; maximum females estimate = 9.19 +/- 1.93 SE; *p*-value = 1.27 x 10^{-5}; Fig. 4B) -- although this increase was not linear. The mean number of grooming partners reached an asymptote between 2-3 grooming partners in large units (maximum females^2 estimate = -4.83 +/- 1.51 SE; *p*-value = 1.59 x 10^{-3}; Fig. 4B). Therefore, while females in small or mid-sized units groomed with more than half (and, in some cases, all) of the females in their unit, females in large units groomed, on average, with fewer than half of the females available. In addition, because grooming time did not increase even in these large units, females in larger units spent less time socializing with each partner than females in smaller units.

These changes in individual grooming behavior suggest that larger units could be more vulnerable to fission than smaller units. We tested this using a social network analysis approach: the network average clustering coefficient, which reflects how interconnected members of a group are to one another (e.g., whether an individual's social partners are also partners). The clustering coefficient represents group cohesion, with lower values reflecting a less cohesive group. We found that the clustering coefficient decreased as unit size increased (maximum females estimate = -0.93 +/- 0.41 SE; *p*-value = 3.39 x 10^{-2}; Fig. 4C). In small units, all individuals were likely to be connected with all other unit members, while in large groups individuals were more fractured.

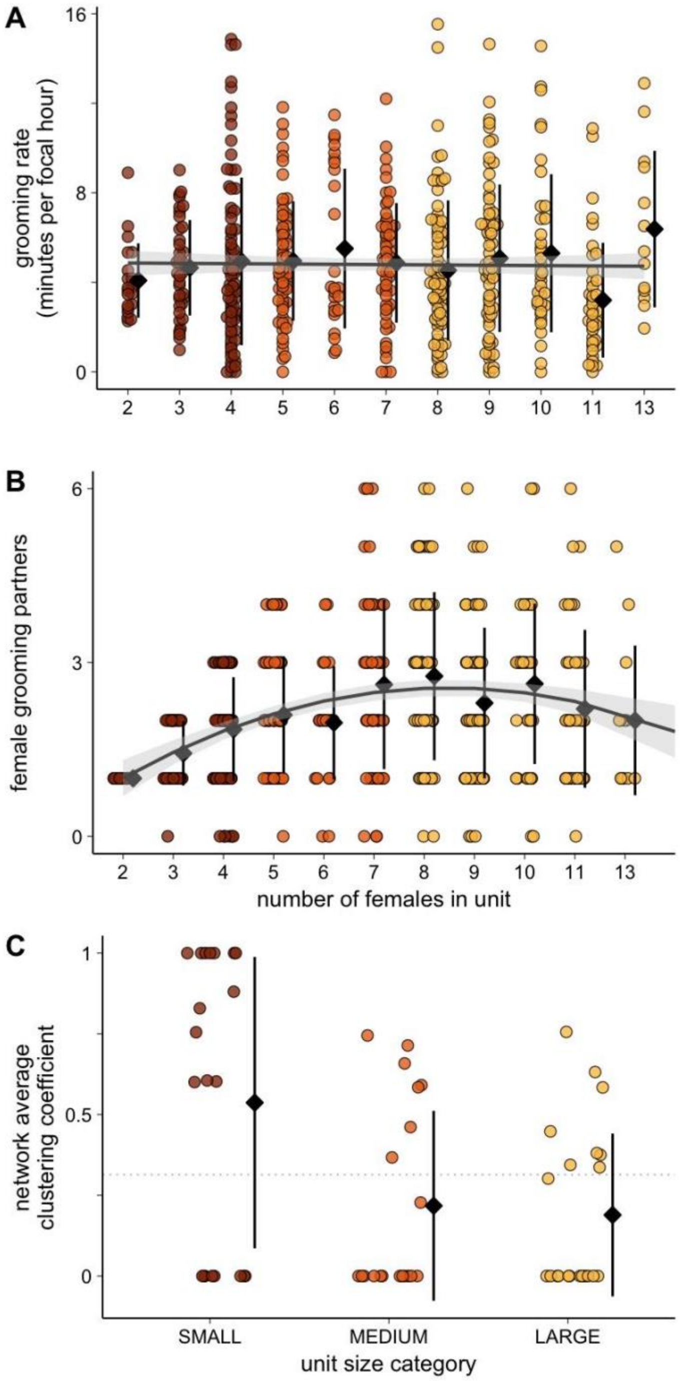

Fig. 4. Larger units are less socially cohesive than smaller units. (A) Total minutes spent grooming by focal hour for each female-year (n = 557 female-years). The x-axis reflects the maximum number of females in at unit across the year. **(B)** Number of female grooming partners a focal female was observed to groom with for each year. **(C)** Network average clustering coefficient by unit size category. Each circle indicates the mean clustering coefficient for a unit at a specific size category for each year. For all figures, black diamonds reflect the mean each unit-size category, with bars representing the standard deviation.

We hypothesize that when fitness costs accrue in large units, females respond by focusing their interactions on a subset of individuals (*33*). This alters the social network structure of the group, increasing the odds of a fission. Gelada social organization may offer a unique opportunity for fissions to occur: gelada units form larger aggregations (i.e., “bands”: *34*) which may buffer the ecological costs of fissions. In contrast, fusions may be less frequent because leader males of small units limit inter-unit interactions among females. Indeed, the two fusions we observed took place immediately after the disappearance of the leader male in one of the units.

Taken together, the individual fitness costs associated with living in either small or large gelada groups – due to the combined effect of both social and ecological pressures – points to the major role of balancing selection in shaping optimal group size. Our findings and approach pave the way for future investigations of the fitness consequences of demographic variation in primates and other long-lived species for which paired fitness and demographic measures have been difficult to quantify.

Acknowledgements: We would like to thank the Ethiopian Wildlife Conservation Authority (EWCA) and the wardens and staff of the Simien Mountain National Park for their permission and ongoing support for our long-term research project. Additionally, we are grateful to our excellent field team across the years, most especially E. Jejaw, A. Fanta, S. Girmay, J. Jarvey, and M. Gomery for their assistance with field data collection. We also owe thanks to A. Marshall, B. Dantzer, and J. Mitani for providing valuable feedback on data analyses. **Funding:** This work was supported by: National Science Foundation (grant numbers BCS-0715179, IOS-1255974, BCS-1340911); Leakey Foundation; National Institutes of Health (grant number R00-AG051764); National Geographic Society (grant numbers 8100-06, 8989-11); Sigma Xi; Wildlife Conservation Society; and the University of Michigan. **Author contributions:** ETJ, NSM, TJB, and JCB designed the study. ETJ, NSM, AL, TJB, and JCB collected the data. ETJ, NSM, TJB, and JCB analyzed the data. ETJ and NSM wrote the paper, with contributions and edits from all authors. ETJ, NSM, AL, TJB, and JCB provided funding support. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data, code, and materials used in this analysis are available here: https://github.com/GeladaResearchProject/Tinsley_Johnson_et_al_2018_Biorxiv.

References and Notes:

1. C. H. Janson, M. L. Goldsmith, Predicting group size in primates: foraging costs and predation risks. *Behav. Ecol.* **6**, 326–336 (1995).

2. M. Grove, Space, time, and group size: a model of constraints on primate social foraging. *Anim. Behav.* **83**, 411–419 (2012).

3. A. C. Markham, L. R. Gesquiere, Costs and benefits of group living in primates: an energetic perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372** (2017), doi:10.1098/rstb.2016.0239.

4. H. R. Pulliam, On the advantages of flocking. *J. Theor. Biol.* **38**, 419–422 (1973).

5. C. P. Van Schaik, Why are diurnal primates living in groups? *Behaviour*, 120–144 (1983).

6. T. H. Clutton-Brock, *Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems* (University of Chicago Press, 1988).

7. L.-A. Giraldeau, The stable group and the determinants of foraging group size. *The ecology of social behavior*, 33–53 (1988).

8. N. Snyder-Mackler, J. C. Beehner, T. J. Bergman, Defining Higher Levels in the Multilevel Societies of Geladas (Theropithecus gelada). *Int. J. Primatol.* (2012), doi:10.1007/s10764-012-9584-5.

9. E. Tinsley Johnson, N. Snyder-Mackler, A. Lu, T. J. Bergman, J. C. Beehner, Social and ecological drivers of reproductive seasonality in geladas. *Behav. Ecol.* (2018),

doi:10.1093/beheco/ary008.

10. J. C. Beehner, T. J. Bergman, Infant mortality following male takeovers in wild geladas. *Am. J. Primatol.* **70**, 1152–1159 (2008).

11. T. J. Bergman, L. Ho, J. C. Beehner, Chest Color and Social Status in Male Geladas (Theropithecus gelada). *Int. J. Primatol.* **30**, 791–806 (2009).

12. E. K. Roberts, A. Lu, T. J. Bergman, J. C. Beehner, Female Reproductive Parameters in Wild Geladas (Theropithecus gelada). *Int. J. Primatol.* **38**, 1–20 (2017).

13. J. C. Jarvey, B. S. Low, D. J. Pappano, T. J. Bergman, J. C. Beehner, Graminivory and Fallback Foods: Annual Diet Profile of Geladas (Theropithecus gelada) Living in the Simien Mountains National Park, Ethiopia. *Int. J. Primatol.* (2018), doi:10.1007/s10764-018-0018-x.

14. P. J. Fashing, N. Nguyen, V. V. Venkataraman, J. T. Kerby, Gelada feeding ecology in an intact ecosystem at Guassa, Ethiopia: variability over time and implications for theropith and hominin dietary evolution. *Am. J. Phys. Anthropol.* **155**, 1–16 (2014).

15. L. A. Isbell, Contest and scramble competition: patterns of female aggression and ranging behavior among primates. *Behav. Ecol.* **2**, 143–155 (1991).

16. A. Koenig, C. Borries, Feeding competition and infanticide constrain group size in wild hanuman langurs. *Am. J. Primatol.* **57**, 33–34 (2002).

17. A. Treves, C. A. Chapman, Conspecific threat, predation avoidance, and resource defense: implications for grouping in langurs. *Behav. Ecol. Sociobiol.* **39**, 43–53 (1996).

18. R. Steenbeek, C. P. van Schaik, Competition and group size in Thomass langurs (Presbytis thomasi): the folivore paradox revisited. *Behav. Ecol. Sociobiol.* **49**, 100–110 (2001).

19. J. A. Teichroeb, E. C. Wikberg, I. Bădescu, L. J. Macdonald, P. Sicotte, Infanticide risk and male quality influence optimal group composition for Colobus vellerosus. *Behav. Ecol.* **23**, 1348–1359 (2012).

20. T. V. Snaith, C. A. Chapman, Primate group size and interpreting socioecological models: Do folivores really play by different rules? *Evolutionary Anthropology: Issues, News, and Reviews*. **16**, 94–106 (2007).

21. R. I. M. Dunbar, P. Dunbar, Social dynamics of gelada baboons. *Anim. Behav.* **22**, 203–& (1974).

22. E. K. Roberts, A. Lu, T. J. Bergman, J. C. Beehner, A Bruce effect in wild geladas. *Science*. **335**, 1222–1225 (2012).

23. E. H. M. Sterck, D. P. Watts, C. P. van Schaik, The evolution of female social relationships in nonhuman primates. *Behav. Ecol. Sociobiol.* **41**, 291–309 (1997).

24. C. A. Chapman, Ecological constraints on group size in three species of neotropical primates. *Folia Primatol.* . **55**, 1–9 (1990).

25. M. Emery Thompson, Energetics of feeding, social behavior, and life history in non-human primates. *Horm. Behav.* **91**, 84–96 (2017).

26. J. C. Beehner, C. McCann, Seasonal and altitudinal effects on glucocorticoid metabolites in a wild primate (Theropithecus gelada). *Physiol. Behav.* **95**, 508–514 (2008).

27. L. A. D. Campbell, P. J. Tkaczynski, J. Lehmann, M. Mouna, B. Majolo, Social thermoregulation as a potential mechanism linking sociality and fitness: Barbary macaques with more social partners form larger huddles. *Sci. Rep.* **8**, 6074 (2018).

28. C. Sueur, O. Petit, J. L. Deneubourg, Short-term group fission processes in macaques: a social networking approach. *J. Exp. Biol.* **213**, 1338–1346 (2010).

29. C. Sueur, J.-L. L. Deneubourg, O. Petit, I. D. Couzin, Group size, grooming and fission in primates: a modeling approach based on group structure. *J. Theor. Biol.* **273**, 156–166 (2011).

30. R. C. Van Horn, J. C. Buchan, J. Altmann, S. C. Alberts, Divided destinies: group choice by female savannah baboons during social group fission. *Behav. Ecol. Sociobiol.* **61**, 1823–1837 (2007).

31. N. F. Koyama, Matrilineal Cohesion and Social Networks in Macaca fuscata. *Int. J. Primatol.* **24**, 797–811 (2003).

32. R. I. M. Dunbar, FUNCTIONAL-SIGNIFICANCE OF SOCIAL GROOMING IN PRIMATES. *Folia Primatol.* . **57**, 121–131 (1991).

33. A. L. Engh *et al.*, Behavioural and hormonal responses to predation in female chacma baboons (Papio hamadryas ursinus). *Proc. Biol. Sci.* **273**, 707–712 (2006).

34. N. Snyder-Mackler, J. C. Beehner, T. J. Bergman, Defining Higher Levels in the Multilevel Societies of Geladas (Theropithecus gelada). *Int. J. Primatol.* (2012), doi:10.1007/s10764-012-9584-5.

35. J. Altmann, Observational study of behavior: sampling methods. *Behaviour*. **49**, 227–267 (1974).

36. D. Bates *et al.*, Lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7). 2014 (2015).

37. R. C. Team, R: A language and environment for statistical computing. R Core Team, Vienna, Austria (2016).

38. J. C. Beehner, P. L. Whitten, Modifications of a field method for fecal steroid analysis in baboons. *Physiol. Behav.* **82**, 269–277 (2004).

39. S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994).

40. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal, Complex Systems*. **1695**, 1–9 (2006).

41. A. Barrat, M. Barthélemy, A. Vespignani, Modeling the evolution of weighted networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **70**, 066149 (2004).

42. D. J. Watts, S. H. Strogatz, Collective dynamics of “small-world”networks. *Nature*.

393, 440 (1998).

43. H. Wickham, ggplot2. *WIREs Comp Stat*. **3**, 180–185 (2011).

SUPPLEMENTARY CONTENT

Materials and Methods

Figures S1

Tables S1-S11 (all Tables will be hosted online in a single *.xls file)

AUTHOR CONTRIBUTIONS

ETJ, NSM, TJB, and JCB designed the study. ETJ, NSM, AL, TJB, and JCB collected the data. ETJ, NSM, TJB, and JCB analyzed the data. ETJ and NSM wrote the paper, with contributions and edits from all authors. ETJ, NSM, AL, TJB, and JCB provided funding support.

MATERIALS AND METHODS

Text S1. Study site and subjects

The data for this study derive from 9 years of observation (2006-2014) on a population of wild geladas living in the Simien Mountains National Park, in northern Ethiopia (13°13.5’ N latitude). The Simien Mountains Gelada Research Project (SMGRP, formerly the University of Michigan Gelada Research Project) has collected behavioral, demographic, genetic, and hormonal data from individuals since Jan 2006. All gelada subjects are habituated to human observers on foot and are individually recognizable. Daily precipitation was measured using a rain gauge, while daily maximum and minimum temperatures were measured using a digital thermometer placed in a permanently shaded area. We used longitudinal data from 189 adult females in 33 reproductive units (20 original “founding” units plus 13 “daughter” units resulting from fissions). All adult females had known or estimated birth dates from which we calculated age. The mean age at the mid-point of the study (2010) for all females was 12.47 +/- 5.27 SD years; overall range = 4.75 - 27.83 years. Estimated birth dates were calculated based on known dates for other reproductive events (e.g., maturation, first birth, age of oldest offspring, or number of known offspring).

We used 15-min focal animal samples (*35*) to record all social behaviors involving adult females, noting behavioral states (feeding, resting, moving, socializing) and grooming behavior, including total time spent grooming and the identities of grooming partners. This dataset represents 1845.5 hours of focal observation (mean = 4.23 +/- 3.57 SD focal hours per female-year).

Text S2. Unit size

The identities of all individuals present in a unit were recorded each day the unit was seen. For each month of the study period, we recorded the total number of adult females in each unit; where changes in unit size occurred (i.e., due to adult female deaths or subadult female maturations), we used the maximum number of adult females in a unit in a given month. We focus on adult females because we have longitudinal records of the number of adults in each unit for the entire study period but only started recording the total number of individuals (including juveniles and infants) in 2012. However, when we compared the maximum number of adult females in a unit to the total number of individuals in a unit for the subset of data where we have both (2012-2014), we found that both values were highly and significantly correlated (Pearson correlation coefficient = 0.91, p-value = 4.5 x 10^{-218}; Fig. S1A). We also calculated the maximum number of adult males in each unit for each month, which was not highly correlated with the maximum number of females (Pearson correlation coefficient = 0.08, p-value = 4.5 x 10^{-2}; Fig. S1B).

Changes in the number of adult females in a unit were either due to female maturations or to deaths. Maturations were recorded as the first observation of a sex skin swelling (details are outlined in *12*). Deaths were recorded as the first day an individual was no longer observed with a unit when that individual was consistently absent for three consecutive encounters with that unit (and not observed in a different unit, as in the case of transfers or fissions).

Dates of fissions, fusions, and female transfers were assigned to the first day unit females were no longer observed together and subsequently observed either in a separate daughter unit with a new leader male (in the case of fissions), together with non-unit females and a new leader male (in the case of fusions), or associating with a different unit and new leader male (in the case of female dispersals). In all cases, we immediately identified known females in daughter units or new units following their disappearance from their natal unit.

Unless noted, all statistical models included the number of females as a continuous predictor variable. However, for visualization purposes, we also categorized unit size into small, medium (i.e., mid-sized), and large units based on the observed range in variation in sizes. Specifically, cut-offs were determined by calculating the tertiles of the observed monthly distribution of sizes: “small” indicates units of less than or equal to 4 adult females; “mid-sized” indicates units between 5 and 7 adult females; “large” indicates units of 8 or more adult females.

Text S3. Reproductive performance

For the majority of infants born during the study period (n = 243 out of 272 total), the date of birth was known within a day or two. For those infants where we did not observe their day of birth (n = 29), we were able to assign the date of birth within 1 month of the actual birth date based on established criteria (i.e., the size of the infant, infant motor skills, presence/absence of the umbilical cord; for more details, see *9*). From these birth dates we were able to calculate conception dates based on the mean gestation length (n = 183 days; for more details, see: *12*).

We used a binary scoring system to assign female reproductive performance. Females were “successful” when their infant survived to 1.5 years of age, which is the mean age of weaning in this population (regardless of actual weaning date). Successful females received 1 point for each month of pregnancy and lactation leading up to the designated end point of 1.5 years (6 months gestation + 18 months of lactation = 24 successful months). If the infant died before reaching 1.5 years of age, no points were awarded to the months of gestation/lactation prior to the infant death.

The disappearance of any infant prior to weaning was assumed to be a case of infant mortality. The cause of mortality was assigned based on the following characteristics: if the infant's mother died at the same time, the cause of death was recorded as "maternal death." If the mother did not die at the same time, and the infant death occurred within 9 months of a takeover, the cause of death was recorded as "infanticide" (*9*, *10*). All other causes of infant deaths were recorded as "unknown."

To assess the effects of unit size on female reproductive performance, we constructed a binomial generalized linear mixed-effects model (GLMM) using the *lme4* package (version 1.1-12: *36*) in R (version 3.3.2: *37*). The dependent binary variable was the monthly success variable for each female-month. We modeled this outcome variable as a function of the following predictors: female age (both the linear and quadratic term, to control for the known effects of female age on reproductive output), maximum number of males in the unit that month (as the number of adult males in a unit varies independently from the number of adult females), and maximum number of females in the unit that month (both the linear and the quadratic term). We controlled for the repeated measures of individual identity, month, and year as random effects.

To account for the potential effect of pseudoreplication in our approach we also binned our success variable across two different time periods (3 months and 6 months) and constructed two new binomial GLMMs. For each of these models, the outcome variable was the number of successful months for each female over the given period out of the number of unsuccessful months over the given period. We included individual identity and the time period (i.e., the specific 3- or 6-month window) as random effects.

The results of this binning approach (summarized in Table S2 and S3) recapitulated those of our original month-by-month analysis.

Text S4. Adult female death rate

For each study year, we calculated the total time (in female-years) each unit was observed at a particular size and summed the number of female deaths observed in that unit at each size. The death rate was calculated for each unit-size as the number of adult female deaths out of the number of individual females observed in that unit size for each year, and was modeled as a function of unit size, controlling for time observed (in female-years) at each size.

To assess whether the rate at which adult females died varied by unit size, we first calculated the number of deaths observed for each unit-year and the size of the unit at the time of each death. For each year in our analysis, we summed the total time (in female-years) that we observed each unit at a specific size and matched the time observed at a given size with the total deaths observed that year at that size. We constructed a binomial GLMM using the *lme4* package. The dependent binomial variable was the number of females that died out of the number of females that survived during the time a unit was observed at a specific size. We modeled this outcome variable as a function of the following predictors: unit size (the maximum number of females in the unit, including both the linear and the quadratic term) and the mean age of all unit females observed (to control for the increasing risk of death with age). We controlled for the repeated measures of unit and year as random effects, and to control for observation time we offset our outcome variable by the number of female-years a unit was observed at that specific size in a given year.

Text S5. Takeover rate

We recorded the dates of all observed male takeovers (*n* = 72) of known reproductive units (following *9*) as well as the number of mature females in the unit at the time of takeover.

To assess how infanticide risk varied with unit size, we first considered takeover frequency, as infanticides almost exclusively occur within the context of takeovers. We calculated the total number of takeovers observed according to the size of the unit (i.e., number of adult females) at the time of takeover. We modeled this number as the dependent variable in a Poisson GLMM, offset by the total observation time (unit-years) of all units at each unit size. We controlled for the repeated effects of unit by including unit as a random intercept. Finally, we included two fixed effect variables: the maximum number of adult females in the unit and the average number of males in each size of unit.

Text S6. Infanticide rate

Next, we considered whether infanticide rates varied by unit size. We first calculated the number of infant births for each unit-year and the size of the unit at the time of each birth (total *N* births = 269). Out of these births, we also calculated the number of infants that subsequently died due to infanticide before reaching 1.5 years of age, i.e., the mean age at weaning in this population (*12*; total *N* infanticides = 33). For each year in our analysis, we summed the total time (in female-years) we observed each unit at a specific size and matched the time observed at a given size with the number of births and the subsequent number of infanticides observed that year at that unit size. We constructed a binomial GLMM using the *lme4* package. The dependent binomial variable was the number of births that resulted in an infanticide out of the number of births that did not result in an infanticide. We modeled this outcome variable as a function of unit size at birth (the maximum number of females in the unit, including both the linear and the quadratic term). We controlled for the repeated measures of unit and year as random effects, and to control for observation time, we offset our outcome variable by the number of female-years a unit was observed at that specific size in a given year.

Text S7. Feeding time

Feeding time was recorded by noting the broad behavioral state (feeding, moving, resting, or social) continuously throughout each 15-min focal. For each month, we calculated total time spent feeding (minutes per focal hour) and limited our analysis to females that had 3 or more separate focals for that month (>= 0.75 focal hours: *n* = 132 adult females over 58 months of observation).

To investigate the relationship between feeding time and unit size, we constructed a series of linear mixed-effects models (LMM) using the *lme4* package. Our dependent variable was the total minutes an individual spent feeding in a given month offset by the time observed that month (focal minutes). We considered seven variables in total. First, we included factors related to the size of the unit: maximum number of males in the unit that month, maximum number of females in the unit that month (including either the linear term or the quadratic term for number of females), and the unit size category (small, medium, large) that matched the size of the unit. Second, we included factors related to the seasonal variation in temperature and food availability: maximum temperature (mean maximum temperature across the previous 30 days) or rain (cumulative rain over the previous 90 days: *9*, *13*). For maximum temperature, we included either the linear or the quadratic term. Because temperature and rain were correlated (Pearson's correlation: $r = -0.696$, $p < 2.2 \times 10^{-16}$), we ran two sets of LMMs, one that included temperature (or temperature-squared) and one that included rain. Finally, we included factors related to individual condition, which could impact energetic demand: age (maximum female age for the month) and reproductive state (cycling, lactating, or pregnant).

We compared all candidate models using AIC and present here the results of the top two models: (i) maximum females^2, maximum number of males, maximum temperature^2, maximum female age, and reproductive state (see Table S7) and (ii) the unit size category, maximum number of males, maximum temperature^2, maximum female age, and reproductive state (see Table S8).

Text S8. Glucocorticoid metabolites

We collected fecal samples from 148 known adult females between 2006 and 2014 (n = 3835 hormone samples; mean = 26 samples per female; range: 1-150 samples per female). Fecal samples were collected using noninvasive methods developed by the SMGRP for hormone extraction and preservation under field conditions (*38*). All samples were assayed for glucocorticoid metabolites (GCMs) using reagents from the ImmuChem^{TM} double antibody corticosterone ^{125}I RIA kit (MP Biomedicals, LLC, Orangeburg, NY). (For more details on hormone extraction and assay methods see: *9*).

To assess the effect of unit size on GCMs, we log-transformed GCM values to approximate a normal distribution, and then modeled logGCMs (in a LMM using *lme4*) as a function of the following predictors. First, we included a variety of predictors previously shown to affect female GCMs (*9*) due to (1) seasonal variation in temperature, (2) individual characteristics, and (3) social events. Specifically, we included mean maximum and mean minimum temperature over the 30 days prior to sample collection, female age (which we included as both a linear and quadratic term), and reproductive state (pregnant, cycling, or lactating), as well as the interactions between age and age-squared and reproductive state. We also included a categorical predictor reflecting whether the sample had been collected in the 30 days following a takeover, and the interaction between takeover and reproductive state.

Finally, we included two predictors reflecting unit size: maximum number of unit females (both the linear and quadratic term) and maximum number of unit males. We controlled for the repeated effects of individual identity, month, unit, and year (for results, see Table S9).

Text S9. Grooming time, partners, and network average clustering coefficient

All grooming interactions, including the duration and identities of grooming partners, were also recorded. For each year, we first calculated *grooming time* (minutes of grooming, both given and received, per focal hour) for all co-resident females (i.e., for all potential grooming partners). We also summed the number of adult female grooming partners each focal was observed to groom with each year to calculate their *total number of grooming partners*. Finally, we used social network analysis to assess the global properties of the social network of each unit (*39*). Specifically, we used the package *igraph* (version 1.0.0, *40*, in R to first calculate the individual clustering coefficient, or weighted transitivity, for each female in a unit based on dyadic grooming rates (using the function *transitivity*, *41*; at the local, or individual level, the clustering coefficient reflects how embedded an individual is in a social network; i.e., how well-connected they are as well as how well-connected their associates are). We limited this analysis to units with 3 or more adult females, because the clustering coefficient is based on the number of “closed” triangles in a network (i.e., how many triplets each share ties). To assess the interconnectivity of the unit at the global level, or the *network average clustering coefficient*, we calculated the mean individual clustering coefficient for all adult females in the same unit each year (*42*). The network average clustering coefficient indicates the degree to which all individuals in a unit cluster together.

To assess the effect of unit size on female social behavior, we constructed three sets of LMMs using the *lme4* package. The outcome variables for each set of models were: (1) *grooming time*: total yearly minutes grooming per focal hour; (2) *grooming partners*: total yearly adult female grooming partners per focal female; and (3) *network average clustering coefficient*. We modeled each outcome variable as a function of the following predictors: the maximum number of adult females in the unit (including the linear and the quadratic terms), the maximum number of males in the unit, and (for the first two individual-level measures) the focal female's age at the end of the year. We controlled for the repeated effects of individual identity (for *grooming time* and *grooming partners*), unit, and year, and limited our analysis to females that had at least one hour of focal observation during the year (four 15-min focals). Each model set included the intercept-only null model, univariate models that considered only a single fixed effect, and multivariate models that considered all combinations of fixed effects. Model fits were compared using AIC and the results of the top model are reported here (see Tables S10-S12).

Figures were constructed using *ggplot2* (*43*).

FIGURES S1.

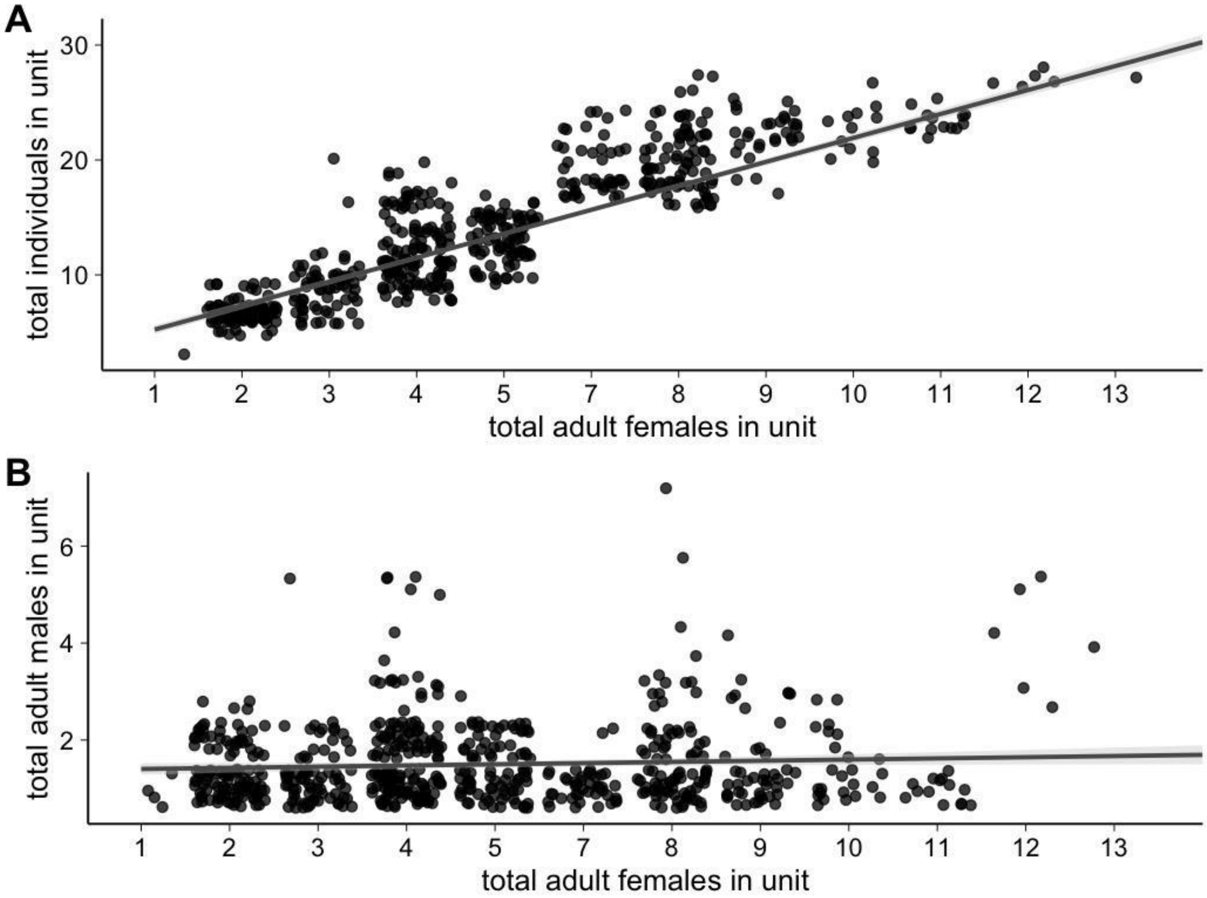

Fig. S1. The number of adult females in a unit is highly correlated with the total number of individuals in a unit. (A) Total adult females in a unit (x-axis) by the total individuals (y-axis; including juveniles and infants) in a unit for each unit-month. **(B)** Total adult females in a unit (x-axis) by the total adult males (y-axis) in a unit for each unit-month.

TABLES S1-S12.

Table S1. Results from reproductive performance model (1-month intervals)

Predictor	Estimate	Standard Error	z value	*p*-value
Female age	-87.70	10.02	-8.75	*** 2.1×10^{-18}
Female age^2	-35.86	4.34	-8.27	*** 1.3×10^{-16}
Maximum males	-0.40	0.04	-11.35	*** 7.1×10^{-30}
Maximum females	-4.03	4.27	-0.94	3.4×10^{-1}
Maximum females^2	-11.22	3.30	-3.40	*** 6.7×10^{-4}

Table S2. Results from reproductive performance model (3-month intervals)

Predictor	Estimate	Standard Error	z value	*p*-value
Female age	-52.48	7.28	-7.21	*** 5.6×10^{-13}
Female age^2	-21.61	2.74	-7.89	*** 3.4×10^{-15}
Maximum males	-0.43	0.03	-12.64	*** 1.3×10^{-36}
Maximum females	-2.27	2.64	-0.86	3.9×10^{-1}
Maximum females^2	-4.81	2.09	-2.31	* 2.1×10^{-2}

Table S3. Results from reproductive performance model (6-month intervals)

Predictor	Estimate	Standard Error	z value	*p*-value
Female age	-34.56	5.87	-5.89	*** 3.8×10^{-9}
Female age^2	-16.06	2.05	-7.84	*** 4.6×10^{-15}
Maximum males	-0.41	0.03	-13.33	*** 1.4×10^{-40}
Maximum females	0.73	1.99	0.37	7.1×10^{-1}
Maximum females^2	-4.00	1.58	-2.54	* 1.1×10^{-2}

Table S4. Results from female death rate model

Predictor	Estimate	Standard Error	z-value	*p*-value
Maximum females	-23.10	3.12	-7.41	*** 1.3×10^{-13}
Maximum females^2	13.52	3.01	4.49	*** 7.1×10^{-6}
Female age	0.20	0.07	2.63	** 8.5×10^{-3}

620

Table S5. Results from takeover rate model

Predictor	Estimate	Standard Error	z-value	*p*-value
Maximum females	0.51	0.20	2.49	* 1.3×10^{-2}
Maximum males	0.64	0.18	3.59	*** 3.4×10^{-4}

621

Table S6. Results from infanticide rate model

Predictor	Estimate	Standard Error	z-value	*p*-value
Maximum females	-18.35	6.51	-2.82	** 4.8×10^{-3}
Maximum females^2	24.40	5.85	4.17	*** 3.0×10^{-5}

622

Table S7. Results from top model for monthly feeding time

Predictor	Estimate	Standard Error	t-value	*p*-value
Maximum females^2	2.30	0.73	3.16	** 3.2×10^{-3}
Maximum males	-2.12	0.54	-3.92	*** 9.6×10^{-5}
Maximum temperature^2	3.06	1.40	2.18	* 3.1×10^{-2}
Female age	-1.01	0.45	-2.23	* 2.7×10^{-2}
Reproductive state: lactating	1.14	1.05	1.08	2.8×10^{-1}
Reproductive state: pregnant	0.27	1.30	0.21	8.3×10^{-1}

Table S8. Results from top model for monthly feeding time (categorical predictor)

Predictor	Estimate	Standard Error	t-value	*p*-value
Unit category: Medium	-1.89	1.54	-1.23	2.2×10^{-1}
Unit category: Large	4.01	1.92	2.09	* 4.2×10^{-2}
Maximum males	-2.12	0.54	-3.92	*** 3.8×10^{-5}
Maximum temperature^2	3.06	1.40	2.18	* 3.2×10^{-2}
Female age	-1.01	0.45	-2.23	* 2.2×10^{-2}
Reproductive state: lactating	1.15	1.05	1.08	2.7×10^{-1}
Reproductive state: pregnant	0.15	1.30	0.12	9.1×10^{-1}

Table S9. Results from glucocorticoids model

Predictor	Estimate	Standard Error	t-value	*p*-value
Maximum females	-0.91	0.61	-1.50	1.4 x 10^{-1}
Maximum females^2	0.67	0.44	1.53	1.3 x 10^{-1}
Maximum males	0.03	0.01	4.73	*** 2.5 x 10^{-6}
Maximum temperature	-0.06	0.01	-5.24	*** 2.5 x 10^{-7}
Minimum temperature	-0.005	0.01	-0.59	5.5 x 10^{-1}
Takeover (Y)	0.06	0.02	3.16	** 1.6 x 10^{-3}
Reproductive state: lactating	0.08	0.01	6.29	*** 3.9 x 10^{-10}
Reproductive state: pregnant	0.11	0.01	8.69	*** < 2.0 x 10^{-16}
Female age	2.27	0.45	4.99	*** 1.2 x 10^{-6}
Female age^2	-1.97	0.40	-4.90	*** 1.5 x 10^{-6}
Takeover (Y) x Reproductive state: lactating	-0.03	0.04	-0.68	5.0 x 10^{-1}
Takeover (Y) x Reproductive state: pregnant	0.25	0.07	3.49	*** 4.9 x 10^{-4}
Reproductive state: lactating x Female age	-1.81	0.70	-2.60	** 9.3 x 10^{-3}
Reproductive state: pregnant x Female age	-2.58	0.70	-3.68	*** 2.4 x 10^{-4}
Reproductive state: lactating x Female age^2	3.354	0.82	4.30	*** 1.8 x 10^{-5}
Reproductive state: pregnant x Female age^2	2.16	0.78	2.78	** 5.6 x 10^{-3}

Table S10. Results from grooming time model

Predictor	Estimate	Standard Error	t-value	*p*-value
Maximum females	1.35	4.43	0.31	7.6×10^{-1}
Maximum females^2	-3.47	3.74	-0.93	3.6×10^{-1}
Maximum males	0.48	0.14	3.44	*** 6.4×10^{-4}
Female age	-0.42	0.16	-2.71	** 7.4×10^{-3}

628

Table S11. Results from number of grooming partners model

Predictor	Estimate	Standard Error	t-value	*p*-value
Maximum females	8.88	1.94	4.58	*** 2.6×10^{-5}
Maximum females^2	-5.17	1.51	-3.42	*** 7.5×10^{-4}
Maximum males	0.08	0.05	1.45	1.5×10^{-1}
Female age	-0.11	0.06	-1.97	. 5.1×10^{-2}

629

Table S12. Results from clustering coefficient model

Predictor	Estimate	Standard Error	t-value	*p*-value
Maximum females	-0.93	0.41	-2.26	* 3.4×10^{-2}
Maximum females^2	0.41	0.37	1.12	2.7×10^{-1}

630