LETTER

https://doi.org/10.1038/s41586-018-0422-6

Fitness benefits and emergent division of labour at the onset of group living

Y. Ulrich[1,2]*, J. Saragosti[1], C. K. Tokita[3], C. E. Tarnita[3] & D. J. C. Kronauer[1]*

The initial fitness benefits of group living are considered to be the greatest hurdle to the evolution of sociality[1], and evolutionary theory predicts that these benefits need to arise at very small group sizes[2]. Such benefits are thought to emerge partly from scaling effects that increase efficiency as group size increases[3–5]. In social insects and other taxa, the benefits of group living have been proposed to stem from division of labour[5–8], which is characterized by between-individual variability and within-individual consistency (specialization) in task performance. However, at the onset of sociality groups were probably small and composed of similar individuals with potentially redundant—rather than complementary—function[1]. Self-organization theory suggests that division of labour can emerge even in relatively small, simple groups[9,10]. However, empirical data on the effects of group size on division of labour and on fitness remain equivocal[6]. Here we use long-term automated behavioural tracking in clonal ant colonies, combined with mathematical modelling, to show that increases in the size of social groups can generate division of labour among extremely similar workers, in groups as small as six individuals. These early effects on behaviour were associated with large increases in homeostasis—the maintenance of stable conditions in the colony[11]—and per capita fitness. Our model suggests that increases in homeostasis are primarily driven by increases in group size itself, and to a smaller extent by a higher division of labour. Our results indicate that division of labour, increased homeostasis and higher fitness can emerge naturally in social groups that are small and homogeneous, and that scaling effects associated with increasing group size can thus promote social cohesion at the incipient stages of group living.

Quantifying the effects of group size on fitness and division of labour (DOL) requires the ability to precisely manipulate group size, monitor individual behaviour within groups and accurately measure fitness in controlled conditions. Crucially, group size must be controlled independently from factors such as colony genetic or age structure, which often co-vary with group size and which can affect fitness and DOL[11,12]. To overcome these challenges, we use the clonal raider ant *Ooceraea biroi*, which combines the rich social biology of ants with unprecedented experimental amenability. This species displays an unusually simple social organization: colonies have no queens, and consist of genetically identical, monomorphic, totipotent workers that reproduce clonally and synchronously, and emerge in discrete age cohorts[13]. This provides maximal experimental control over group size and the genetic and demographic structure of colonies. Synchronized reproduction drives stereotypical colony cycles, in which colonies alternate between reproductive and brood-care phases, corresponding to the absence and presence of larvae. During the reproductive phase, all ants remain inside the nest and lay eggs. During the brood-care phase, the ants attend to the larvae inside the nest but also leave the nest—for example, to forage and to dispose of waste[14].

We monitored the behaviour and fitness of colonies containing between 1 and 16 ants matched for genotype and age over at least one colony cycle (Methods). Thus, within- and between-colony variation in genotype and age were minimal and could be ruled out as sources of variation in behaviour and fitness. Workload was standardized using a fixed initial larvae-to-workers ratio (1:1) across group sizes. The experiment was performed with 112 colonies from two clonal genotypes, A and B (previously labelled as MLL1 and MLL4[15]). Experiments started in the brood-care phase with workers and young larvae, and ended when all larvae in all colonies had either developed into adults or died. Colonies were kept in Petri dishes with no brood chamber (Fig. 1a); the ants freely chose a location to place their brood pile (henceforth, 'the nest'). We analysed behaviour using custom automated image acquisition (7–9 frames per hour over 39–41 days) and analysis tools (Fig. 1a, Extended Data Fig. 1, Methods).

Because work in insect societies is spatially organized[16,17] (for example, foraging and waste disposal occur away from the nest, whereas nursing occurs at the nest), individual behaviour can be described in terms of spatial location[18]. This is commonly done by assigning individuals to discrete behavioural groups on the basis of manually acquired spatial data[13]. Although the acquisition of spatial data has greatly improved with automated behavioural tracking[18–20], individuals are often still clustered into discrete behavioural groups[18]. However, in many systems—especially those without morphological castes—individual behaviour is continuously distributed[21,22]. We therefore analysed behaviour non-parametrically from continuous spatial data, avoiding assumptions about the statistical distribution of individual behaviour. The spatial distribution of each ant was measured as the two-dimensional root-mean-square deviation (r.m.s.d.) of its x and y coordinates; that is, the spread of these coordinates around their centre of mass, throughout the brood-care phase (Fig. 1b, Extended Data Fig. 2a, Methods). The r.m.s.d. of an ant captures its tendency to explore the arena—that is, its lack of spatial fidelity (Fig. 1c)—and strongly correlates with its mean distance to the nest (Extended Data Fig. 3). The r.m.s.d. value is therefore a biologically meaningful metric that reflects the propensity to perform tasks away from the nest (for example, foraging) rather than at the nest (for example, nursing). In fact, the mean r.m.s.d. of a colony reflects its foraging activity: it increases when nutritional demand is elevated by increasing the larvae-to-workers ratio (Extended Data Fig. 4a, Supplementary Methods). As expected in this system, individuals varied in r.m.s.d., but did not cluster into discrete behavioural groups (Fig. 1b).

If DOL increases with group size, larger colonies are expected to show (1) higher behavioural variation between colony members and (2) higher individual behavioural consistency, or specialization, over time. Although not always independent from each other, these measures reflect distinct facets of DOL. Behavioural variation, computed as the standard deviation across r.m.s.d. values of ants from the same colony, increased with group size (Fig. 1d, Extended Data Fig. 5a), with small colonies (sizes 2–8) displaying less behavioural variation than larger colonies (sizes 12–16) (colony or group size refers to the number of individual workers and larvae added at the beginning of the experiment; for example, a colony of size 2 comprised 2 workers and 2 larvae). Short-term specialization was quantified as the r.m.s.d. rank

[1]Laboratory of Social Evolution and Behavior, The Rockefeller University, New York, NY, USA. [2]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland. [3]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. *e-mail: yuko.ulrich@gmail.com; dkronauer@rockefeller.edu

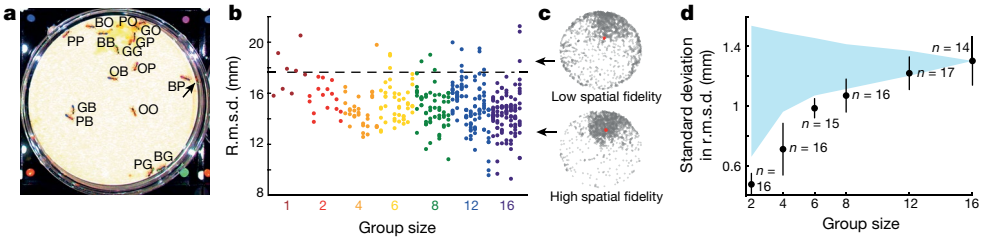


Fig. 1 | Behavioural variation as a function of group size. a, Example frame showing the result of automated ant detection: 15 correct colour-tag assignments (B, blue; G, green; O, orange; P, pink), 1 missed ant (arrow). **b**, Individual r.m.s.d. values for workers of genotype A. Ants from the same colony are vertically aligned. For the definition of r.m.s.d., see 'Behavioural data analysis' in Methods. Dashed line denotes the expected r.m.s.d. assuming a uniform distribution of an ant's positions. **c**, Spatial distribution of two ants from the same colony over the brood-care phase. Red, centre of mass. Arrows point to the corresponding ants in **b**. Note that even workers with low spatial fidelity spend most of their time in the nest area. **d**, Behavioural variation increases with group size. Data for genotypes A and B are pooled. Black, standard deviation in r.m.s.d. per colony as a function of group size (mean ± s.e.m.). Blue, 95% confidence intervals under the null hypothesis of no group-size effect on individual behaviour, generated by resampling individuals from colonies of size 16 (Extended Data Fig. 5a).

correlation between consecutive days, averaged over the first brood-care phase (Fig. 2a, Extended Data Fig. 5b–d); this captures the day-to-day behavioural consistency of group members relative to each other. Short-term specialization increased with group size and became significantly different from random at group size 6 (Fig. 2b). Long-term specialization, computed as the correlation between individual mean-r.m.s.d. ranks in the first and second brood-care phases (Extended Data Fig. 5e), was also found in colonies of six or more workers (Fig. 2c–f). Thus, DOL emerged at small group sizes, even in the absence of genetic and age variation, and increased as groups became larger.

We next used mathematical modelling to explore whether fixed response thresholds, the best theoretically studied self-organizing mechanism for DOL[23], could recapitulate these results. Individuals are assumed to respond to two task-related stimuli that reflect colony demand. A response occurs on the basis of innate, fixed thresholds that determine the propensity of individual ants to perform a task given a certain stimulus level. For each task, individual thresholds are drawn from a normal distribution. The lower the stimulus intensity compared to an individual's threshold, the less likely it is to perform the task. The more sensitive this decision is to differences between stimulus intensity and threshold level, the more deterministic the threshold response (Methods). Individuals do not differ in their ability to perform tasks and there are no task-switching costs. When a task is performed, its stimulus level decreases; otherwise, it increases. Thus, across time, individuals divide their effort between tasks in various proportions, recapitulating a behavioural continuum that is similar to the experimental data. As long as there was some threshold variation among individuals, this simple model could robustly produce increased specialization with group size (measured as the slope between specialization values at group sizes 2 and 16) across a large parameter space,

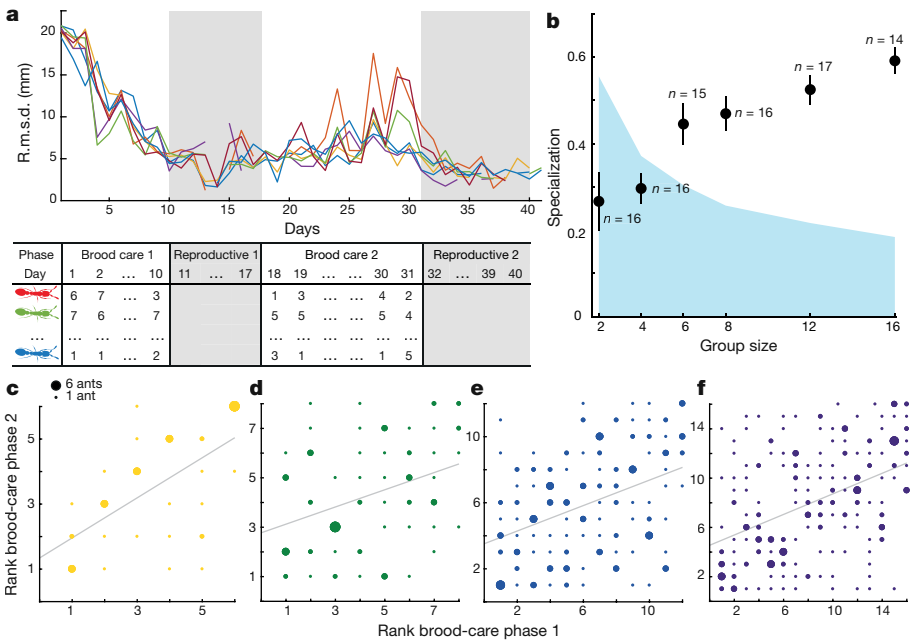


Fig. 2 | Specialization as a function of group size. a, Daily individual r.m.s.d. in one colony (size 8, genotype B). The matrix shows a subset of the corresponding daily r.m.s.d. ranks. **b**, Specialization (mean ± s.e.m.) increases with group size. Black, mean day-to-day r.m.s.d.–rank correlation coefficients in the first brood-care phase as a function of group size. Positive values indicate a tendency for workers to maintain their behavioural rank across days, and 0 indicates that ranks are random. Blue, 95% confidence intervals generated by randomizing daily ranks (shuffling values along the columns of the matrix in **a**). **c**–**f**, Specialization persists across cycles in colonies of sizes 6–16. Grey lines, least-squares fit. Spearman correlation between individual ranks over successive brood-care phases in colony size 6 (**c**) (r(degrees of freedom: 42) = 0.62, $P = 1.38 \times 10^{-5}$, 95% confidence interval (CI) −0.32 to 0.32), colony size 8 (**d**) ($r(75) = 0.35$, $P = 0.002$, 95% CI −0.23 to 0.24), colony size 12 (**e**) ($r(160) = 0.40$, $P = 1.31 \times 10^{-7}$, 95% CI −0.16 to 0.15), and colony size 16 (**f**) ($r(209) = 0.44$, $P = 3.42 \times 10^{-11}$, 95% CI −0.13 to 0.13). Circle diameter is proportional to the number of ants. Colonies of size 4 ($r(32) = 0.08$, $P = 0.68$, 95% CI: −0.42 to 0.42) are not shown. In **b**–**f**, data for genotypes A and B are pooled.

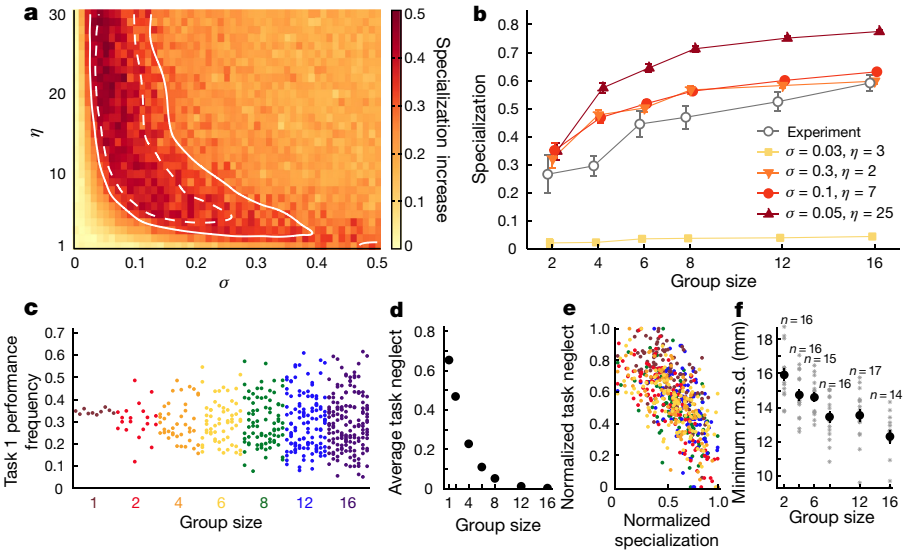


Fig. 3 | Results and predictions of the theoretical model. One hundred replicates were simulated per group size for each parameter combination. **a**, Increase in specialization between group size 2 and 16 (measured as the slope between these values) as a function of threshold stochasticity, η (higher η = a more deterministic threshold), and threshold variation, σ. The region between the solid (lower bound) and dashed (upper bound) white contours encompasses simulated slope values approximated to be within 10% of the experimental slope. **b**, Specialization (mean ± s.e.m.) as a function of group size for different parameter combinations of the theoretical model (colour curves) and for experimental data (grey curve). **c**–**e**, One set of parameters (shown in Extended Data Fig. 6) corresponding to the filled circle symbol in **b**. **c**, Performance frequency of task 1 in simulated colonies of various sizes (10 example replicates shown per group size). Each point represents an ant; ants from the same colony are vertically aligned. **d**, Average task neglect (that is, the proportion of time during a simulation run in which a task went unperformed) across tasks. Points represent the average value (mean ± s.e.m.) across all simulated colonies. **e**, Relationship between specialization and task neglect when controlling for group size. Each point represents one simulated colony; colonies are coloured by group size as in **c**. **f**, Minimum r.m.s.d. (mean ± s.e.m.), an empirical proxy for task neglect, decreases with group size (log-likelihood ratio test: $\chi^2 = 57.79$, $P = 2.92 \times 10^{-14}$). Asterisks represent colony data. Data for genotypes A and B are pooled.

including regions with a close quantitative match to our empirical observations (Fig. 3a, b, Supplementary Methods). Behavioural variation also increased with group size (Fig. 3c, Extended Data Fig. 6a, b). Beyond the group sizes used in our experiments, the model did not predict major further increases in DOL (Extended Data Fig. 6b, c). To explore how DOL might further increase with group size, other mechanisms—for example, direct social interactions[24] or spatial arrangement of tasks[25]—would need to be considered.

The theoretically predicted increase in DOL with group size was robust across different specialization metrics[10,26] (Extended Data Fig. 6d). Of these, the rank correlation metric (Fig. 2b) produced the highest values, which suggests it might be more sensitive to rudimentary types of specialization. However, the other metrics also reveal important insights: for example, task consistency—which quantifies how infrequently individuals switch between tasks (Supplementary Methods)—increased with group size even in the absence of task-switching costs, which suggests that reduced task switching could be an early emergent property of group living.

The theoretical analysis further revealed that increasing group size leads to increased homeostasis by (i) stabilizing stimuli intensities and task performance frequencies over time (Extended Data Fig. 7a, b), and (ii) decreasing task neglect—that is, instances in which tasks are not performed by any ant (Fig. 3d). These effects were obtained by increasing group size alone, even in the absence of DOL—that is, when there was no threshold variation—as long as some other source of stochasticity reduced the likelihood of behavioural synchronization among individuals (Extended Data Fig. 7a–c, Supplementary Notes). However, when present, DOL further increased homeostasis by enhancing some (Fig. 3e, Extended Data Fig. 7d)—although not all (Extended Data Fig. 7e)—of these effects. Larger colonies were thus more homeostatic than smaller colonies but, at a given size, more-specialized colonies had higher homeostasis. Subsequent analyses of the experimental data revealed dampened temporal fluctuations in colony-level behaviour (Extended Data Fig. 8a), increases in colony-level spatial stability (Extended Data Fig. 8b) and decreases in minimum r.m.s.d. (a proxy for task neglect; Methods, Fig. 3f, Extended Data Fig. 9) with increasing group size, consistent with model predictions. These theoretical and empirical findings point to increases in colony homeostasis with group size via temporally more stable levels of work demand (for example, larval hunger) and more consistent work performance (for example, fewer instances of unattended brood). Because colony homeostasis is considered to be a key determinant of colony performance[11], the above results suggest that colony fitness should increase with group size.

Empirically, increases in group size were indeed associated with steep increases in fitness (Fig. 4a, Extended Data Fig. 2b). Colony growth rate was negative for the smallest colonies but rapidly increased with group size and plateaued at around 1 (indicating a doubling of size) in colonies of sizes 12 and 16 (Fig. 4b), similar to values reported in colonies that are orders-of-magnitude larger[27]. Differences in colony growth were partially due to unexpected effects on brood development: the time to eclosion was 11 days (45%) longer in the smallest colonies compared to the largest colonies (Fig. 4c). This effect could not be recapitulated by varying larvae number alone (Extended Data Fig. 4b) and therefore probably arose from more efficient brood care in larger colonies. Because *O. biroi* colony cycles are controlled by the brood[14,15], the different times to eclosion suggest that small colonies had prolonged cycles. In fact, some large colonies had produced two cohorts of workers by the time small colonies had produced one (Extended Data Fig. 2b).

Control experiments and further analyses confirmed that none of our results were confounded by ant tagging, or by variation in ant density or morphology (Extended Data Fig. 10, Supplementary Methods).

In conclusion, we find that DOL, increased homeostasis and higher fitness can emerge as a function of group size at the incipient stages of group living. Notably, the rudimentary and flexible DOL demonstrated here does not rely on well-known mechanisms such as morphological caste specialization or age polyethism[12], but instead on plastic behavioural responses to the social environment[28–30]. Although our theoretical model shows that specialization alone can increase group

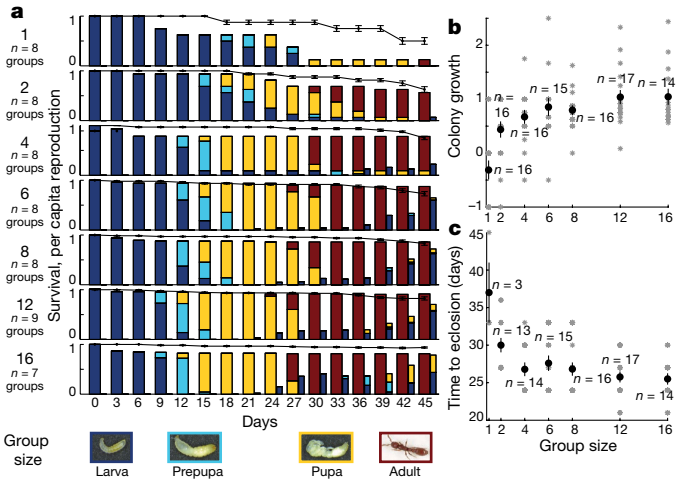


Fig. 4 | Fitness increases with group size. a, The dynamics of brood development as a function of group size in genotype A. Mean proportion of the brood in successive developmental stages (colours); the transition from larvae to pre-pupae marks the end of feeding and the switch from brood-care phase to reproductive phase. Wide and narrow bars indicate first and second brood generations, respectively. Black line, worker survival (mean ± s.e.m.). **b**, Colony growth (mean ± s.e.m.)—calculated as $(W_{end} - W_{start})/W_{start}$, in which W_{end} and W_{start} are the number of live workers at the end and start of the experiment, respectively—increases with group size (log-likelihood ratio test: $\chi^2 = 34.11$, $P = 5.22 \times 10^{-9}$). **c**, The time to eclosion (mean ± s.e.m.) decreases with group size (log-likelihood ratio test: $\chi^2 = 47.92$, $P = 4.44 \times 10^{-12}$). Sample sizes indicate the number of colonies in which at least one larva reached adulthood. In **b**, **c**, asterisks represent colony data, and data for genotypes A and B are pooled.

performance, future work is required to explore whether this direct link between DOL and fitness can be recapitulated experimentally. The scaling effects observed in our experiments and simulations provide a simple mechanism that could, along with other forces, promote social cohesion and provide an evolutionary stepping-stone towards more complex forms of social organization, such as those with morphologically differentiated queen and worker castes.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0422-6.

Received: 23 November 2017; Accepted: 27 June 2018;
Published online 22 August 2018.

1. Queller, D. C. Cooperators since life began. *Q. Rev. Biol.* **72**, 184–188 (1997).
2. Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
3. Berdahl, A., Torney, C. J., Ioannou, C. C., Faria, J. J. & Couzin, I. D. Emergent sensing of complex environments by mobile animal groups. *Science* **339**, 574–576 (2013).
4. Morand-Ferron, J. & Quinn, J. L. Larger groups of passerines are more efficient problem solvers in the wild. *Proc. Natl Acad. Sci. USA* **108**, 15898–15903 (2011).
5. Waters, J. S., Holbrook, C. T., Fewell, J. H. & Harrison, J. F. Allometric scaling of metabolism, growth, and activity in whole colonies of the seed-harvester ant *Pogonomyrmex californicus*. *Am. Nat.* **176**, 501–510 (2010).
6. Dornhaus, A., Powell, S. & Bengston, S. Group size and its effects on collective organization. *Annu. Rev. Entomol.* **57**, 123–141 (2012).
7. Brahma, A., Mandal, S. & Gadagkar, R. Emergence of cooperation and division of labor in the primitively eusocial wasp *Ropalidia marginata*. *Proc. Natl Acad. Sci. USA* **115**, 756–761 (2018).
8. Fewell, J. H. & Harrison, J. F. Scaling of work and energy use in social insect colonies. *Behav. Ecol. Sociobiol.* **70**, 1047–1061 (2016).
9. Jeanson, R., Fewell, J. H., Gorelick, R. & Bertram, S. M. Emergence of increased division of labor as a function of group size. *Behav. Ecol. Sociobiol.* **62**, 289–298 (2007).
10. Gautrais, J., Theraulaz, G., Deneubourg, J. L. & Anderson, C. Emergent polyethism as a consequence of increased colony size in insect societies. *J. Theor. Biol.* **215**, 363–373 (2002).
11. Oldroyd, B. P. & Fewell, J. H. Genetic diversity promotes homeostasis in insect colonies. *Trends Ecol. Evol.* **22**, 408–413 (2007).
12. Jeanson, R. & Weidenmüller, A. Interindividual variability in social insects - proximate causes and ultimate consequences. *Biol. Rev. Camb. Philos. Soc.* **89**, 671–687 (2014).
13. Ravary, F. & Jaisson, P. Absence of individual sterility in thelytokous colonies of the ant *Cerapachys biroi* Forel (Formicidae, Cerapachyinae). *Insectes Soc.* **51**, 67–73 (2004).
14. Ravary, F., Jahyny, B. & Jaisson, P. Brood stimulation controls the phasic reproductive cycle of the parthenogenetic ant *Cerapachys biroi*. *Insectes Soc.* **53**, 20–26 (2006).
15. Oxley, P. R. et al. The genome of the clonal raider ant *Cerapachys biroi*. *Curr. Biol.* **24**, 451–458 (2014).
16. Sendova-Franks, A. B. & Franks, N. R. Spatial relationships within nests of the ant *Leptothorax unifasciatus* (Latr) and their implications for the division of labor. *Anim. Behav.* **50**, 121–136 (1995).
17. Gordon, D. M. Dynamics of task switching in harvester ants. *Anim. Behav.* **38**, 194–204 (1989).
18. Mersch, D. P., Crespi, A. & Keller, L. Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science* **340**, 1090–1093 (2013).
19. Heyman, Y., Shental, N., Brandis, A., Hefetz, A. & Feinerman, O. Ants regulate colony spatial organization using multiple chemical road-signs. *Nat. Commun.* **8**, 15414 (2017).
20. Crall, J. D. et al. Spatial fidelity of workers predicts collective response to disturbance in a social insect. *Nat. Commun.* **9**, 1201 (2018).
21. Weidenmüller, A. The control of nest climate in bumblebee (*Bombus terrestris*) colonies: interindividual variability and self reinforcement in fanning response. *Behav. Ecol.* **15**, 120–128 (2004).
22. Campos, D., Bartumeus, F., Méndez, V., Andrade, J. S. Jr & Espadaler, X. Variability in individual activity bursts improves ant foraging success. *J. R. Soc. Interface* **13**, 20160856 (2016).
23. Bonabeau, E., Theraulaz, G. & Deneubourg, J.-L. Quantitative study of the fixed threshold model for the regulation of division of labour in insect societies. *Proc. R. Soc. Lond. B* **263**, 1565–1569 (1996).
24. Pacala, S. W., Gordon, D. M. & Godfray, H. C. J. Effects of social group size on information transfer and task allocation. *Evol. Ecol.* **10**, 127–165 (1996).
25. Franks, N. R. & Tofts, C. Foraging for work: how tasks allocate workers. *Anim. Behav.* **48**, 470–472 (1994).
26. Gorelick, R., Bertram, S. M., Killeen, P. R. & Fewell, J. H. Normalized mutual entropy in biology: quantifying division of labor. *Am. Nat.* **164**, 677–682 (2004).
27. Teseo, S., Châline, N., Jaisson, P. & Kronauer, D. J. C. Epistasis between adults and larvae underlies caste fate and fitness in a clonal ant. *Nat. Commun.* **5**, 3363 (2014).
28. Crall, J. D. et al. Social context modulates idiosyncrasy of behaviour in the gregarious cockroach *Blaberus discoidalis*. *Anim. Behav.* **111**, 297–305 (2016).
29. Freund, J. et al. Emergence of individuality in genetically identical mice. *Science* **340**, 756–759 (2013).
30. Holbrook, C. T., Kukuk, P. F. & Fewell, J. H. Increased group size promotes task specialization in a normally solitary halictine bee. *Behaviour* **150**, 1449–1466 (2013).

Acknowledgements We thank A. Gal for advice on data analysis, O. Feinerman and M. Liu for contributions to the tracking algorithms, S. Leibler, Z. Frentz, and D. Jordan for helpful discussions. This work was supported by grant 1DP2GM105454-01 from the NIH, a Searle Scholar Award, a Klingenstein-Simons Fellowship Award in the Neurosciences, and a Pew Biomedical Scholar Award to D.J.C.K.; Swiss National Science Foundation Early Postdoc. Mobility (PBEZP3-140156) and Advanced Postdoc.Mobility (P300P3-147900) fellowships, and a Rockefeller University Women & Science fellowship to Y.U.; a Kravis Fellowship to J.S.; the National Science Foundation Graduate Research Fellowship under Grant No. DGE1656466 to C.K.T. This is Clonal Raider Ant Project paper number 8.

Reviewer information *Nature* thanks J. O'Dwyer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions This study was conceived by Y.U. and D.J.C.K. Experiments were designed by Y.U. and D.J.C.K. Tracking hardware and software were developed by J.S. and Y.U. Empirical data were analysed by Y.U. Theoretical modelling was performed by C.K.T. and C.E.T. Computational modelling was performed by C.K.T. Y.U. and D.J.C.K. drafted the manuscript. D.J.C.K. supervised the project. All authors revised the manuscript and approved the final version for publication.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0422-6.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0422-6.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to Y.U. or D.J.C.K.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. Individual ants were assigned to experimental treatments (colonies of different size) randomly. The investigators were not blinded to allocation during experiments and outcome assessment.

Experimental design. Experimental colonies were composed of age-matched, one-cycle-old workers (44 and 34 days old for genotypes A and B, respectively; A colonies have slower cycles than B colonies on average) and 4-day-old larvae in airtight Petri dishes (5 cm in diameter, corresponding to about 25 ant body-lengths) with a plaster of Paris floor. All workers and larvae within an experiment—including replicate colonies of all group sizes—were clonally related and sourced from the same stock colony. All workers within an experiment were also collected from the same cohort and had eclosed within a day of each other (owing to the synchronized reproduction of *O. biroi*). From the time they were collected (1–3 days after eclosion) until the start of the experiment, workers were kept together in a box and allowed to go through a full colony cycle. Thus, all workers within an experiment experienced the same environment as larvae and as adults. However, we cannot exclude the possibility that small differences in individual experience occurred even in this common environment before the start of the experiment. All workers were tagged with colour marks on the thorax and gaster using oil-paint markers (uni Paint Markers PX-20 and PX-21). Experimental colonies contained 1, 2, 4, 6, 8, 12 or 16 workers and a matching number of larvae. This 1:1 larvae-to-workers ratio corresponds to the estimated ratio found in a typical (that is, large and healthy) laboratory stock colony in the brood-care phase. The experiment was conducted using two distinct genotypes, A and B[15]. Between 7 and 9 replicate colonies were used for each group size and genotype, giving a total of 112 colonies. *O. biroi* is myrmecophagous and colonies were fed live pupae of fire ant (*Solenopsis invicta*) minor workers. These prey items are small enough to be transported by a single *O. biroi* worker, so small colonies were not disproportionally penalized by the feeding regime.

The experiments took place in a climate room at 25 °C and 75% relative humidity under constant light (*O. biroi* is blind and its behaviour is not affected by light). Every three days, we cleaned and watered the plaster, added one prey item per live larva at a random location within the Petri dish, and recorded adult survival as well as brood survival and development under a stereomicroscope in all colonies (except for eggs, which cannot be counted without substantially disturbing the colony). The experiments ended when all larvae within an experiment had either developed into adult workers or died. Two colonies (of sizes 6 and 16, genotype B) were excluded from all analyses owing to setup errors (incorrect number of workers or larvae at the beginning of the experiment). Note that although we controlled the number of workers and larvae at the beginning of the experiment, these numbers then changed throughout the experiment as workers died and reproduced, and as the brood died or developed into adults.

Image acquisition and ant detection. Behavioural data were acquired using an automated scan-sampling approach, in which a picture of each colony was taken at regular intervals throughout the experiment. For this purpose, we designed and built a setup comprising 28 webcams (Logitech B910 or C910) and controlled LED lighting. Each webcam acquired images of four colonies, and the position of colonies within the setup was randomized. This resulted in 7,976 and 6,429 frames per colony over 39 and 41 days for genotypes A and B, respectively. The difference in overall frame rate between the two experiments stems in part from the variability in image acquisition speed of the computers used to control the webcams (median interval between frames: 420 s for genotype A, 525 s for genotype B), and in part from an approximately 35-h-long scanning interruption in the genotype B experiment. This interruption occurred in the reproductive phase of most colonies (51 out of 56). Because behavioural analyses were conducted on data collected in the brood-care phase only, this interruption did not affect our results and conclusions. Fitness monitoring outlasted behavioural data acquisition by 6 days in genotype A to allow the last callow workers to eclose.

Within-image variation in lighting and hue was corrected by dividing each frame's RGB values by those of an image of a uniformly grey surface taken with the same camera immediately before the start of the experiment (Extended Data Fig. 1a). After manual selection of the image region corresponding to the plaster arena, a Bayesian classifier was used to assign to each pixel a probability of belonging to each of the following eight colour categories: plaster, ant cuticle, shadow, food and colour tags (pink, orange, blue and green) (Extended Data Fig. 1b). Size and colour-probability thresholds were used to detect candidate regions corresponding to ants carrying colour tags (Extended Data Fig. 1c). Candidate ants were oriented on the basis of the relative position of cuticle and tag colour probability maxima along the main axis of each region (for example, given a candidate ant carrying a blue and a green tag, the blue tag can be assigned to the thorax and the green tag to the gaster if pixels with high cuticle colour probability, corresponding to the ant's head, can be found next to the blue but not the green tag) (Extended Data Fig. 1d). Candidate ants were assigned a final ID using Munkres' variant of the Hungarian assignment algorithm (Extended Data Fig. 1e). Performance of the automated assignments was assessed by comparison with manual assignments for 280 frames selected randomly throughout the first brood-care phase and across colonies in the genotype B experiment. On average, the ant identification algorithms correctly identified 77.1% of the ants that could be manually identified (that is, 22.9% of ants were missed). Of all the automated assignments, 94.4% were correct (that is, 5.6% assigned the wrong ID).

Additionally, we performed manual assignments at a higher frequency (every 10 frames) for one 16-worker colony and verified that individual behavioural traits computed from automated assignments correlated with the same traits computed from manual assignments. Individual behavioural r.m.s.d. values calculated from automated tracking data strongly correlated with the same values calculated from manual tracking (Extended Data Fig. 1f). Software for automated image acquisition and analysis was developed in MATLAB.

Behavioural data analysis. We restricted our behavioural analyses to the brood-care phase because worker locomotion, and thus our ability to detect inter-individual behavioural differences, is markedly reduced during the reproductive phase. For each colony, the brood-care phase started at the beginning of the experiment and ended when all larvae had either reached the non-feeding pre-pupal stage (that is, ejected their meconium) or died. The end of the brood-care phase was scored by visual inspection of images. Note that this definition of the brood-care phase—based solely on the brood developmental stage—is discrete and differs from that of a previous study[31], in which the brood-care phase was characterized using both the development of larvae and the foraging activity of workers.

The spatial distribution of each ant throughout the brood-care phase was quantified as the two-dimensional r.m.s.d.

$$\text{r.m.s.d.} = \sqrt{\frac{\sum_i \left((x_i - \bar{x})^2 + (y_i - \bar{y})^2\right)}{n}}$$

in which x_i and y_i are the coordinates of the focal ant in frame i, $\bar{x}$ and $\bar{y}$ are the coordinates of the centre of mass of the focal ant's overall spatial distribution in the considered time frame, and n is the number of frames in which the focal ant was detected. The r.m.s.d. is bounded between 0 and r, the radius of the Petri dish. Workers that spend a lot of time in the nest with the brood (for example, nursing the larvae) and little time performing outside tasks (foraging or waste disposal) have low r.m.s.d. values, whereas workers that spend comparatively more time away from the brood have higher r.m.s.d. values. In previous studies, workers displaying behaviour corresponding to low or high r.m.s.d. have been labelled 'nurses' and 'foragers', respectively. However, given the apparent continuous distribution of r.m.s.d. values across individuals in this study (Fig. 1b), we chose not to cluster individuals into discrete behavioural 'castes'.

For each colony, mean behaviour was computed as the average of individual r.m.s.d. values, and behavioural variability was computed as the standard deviation of individual r.m.s.d. values. Both metrics were then averaged across replicate colonies for each group size. Artefacts due to sampling effects are of particular concern for any experiment in which variation in group size is an experimental treatment. For this reason, whenever possible we compared group sizes using resampling and randomization approaches in addition to standard statistical tests (Supplementary Notes). To assess the significance of any effect of group size on behaviour while ruling out sampling effects, we simulated colonies of sizes 1 to 12 by randomly sampling 1 to 12 individuals (without replacement) from each colony of size 16 (Extended Data Fig. 5a). Mean behaviour and behavioural variability were calculated for each simulated colony and averaged across replicate colonies of a given size, as described above. This resampling procedure was repeated 1,000 times. Ninety-five per cent confidence intervals were generated for mean behaviour and behavioural variation for each group size separately using the same resampled data, to test which colony size had an observed behaviour that significantly differed from that of colonies of size 16.

To quantify specialization, we introduce a metric appropriate for use in very small colonies and on continuous behavioural data. Existing measures of specialization usually require discrete tasks to be defined and scored, which is generally done by a human observer. Task definition is subjective with respect to the nature and number of defined tasks, and task scoring is susceptible to inter-observer variation. Thus, instead of using existing task-based measures of specialization, our metric is based on continuous behavioural data (r.m.s.d.): colony-level behavioural consistency, or specialization, was defined for each colony as the correlation coefficient between individual r.m.s.d. ranks on consecutive days, averaged over the first brood-care phase (Extended Data Fig. 5b–d). Spearman rank correlations, rather than parametric correlations (for example, Pearson), were used because there was more variation in r.m.s.d. over time than across individuals (Fig. 2a). The timeframe of days was chosen to ensure that individual r.m.s.d. values were calculated on a sufficient number of detections (150–200) for each time interval. These mean rank-correlation coefficients were then compared across colonies of different sizes. This measure can

be used to describe specialization in very small colonies (starting at size 2). To assess significance for each group size, 95% confidence intervals for rank-correlation coefficients were generated by randomizing ranks on each day in each colony 1,000 times, based on the null hypothesis that worker behaviour is uncorrelated across successive days. We then tested whether individual behaviour was consistent over successive brood-care phases. To do this, we selected colonies that had a second brood-care phase (defined by the presence of a new cohort of larvae hatched from eggs laid by the workers during the first reproductive phase) for which at least four days of behavioural data were available. In these colonies, for each brood-care phase, workers were assigned a within-colony rank on the basis of their mean r.m.s.d. across days (Extended Data Fig. 5e). Long-term behavioural consistency was defined as a significant positive correlation between the individual ranks in each brood-care phase. To assess significance, 95% confidence intervals for rank-correlation coefficients were generated by randomizing ranks in each brood-care phase for each colony 1,000 times, based on the null hypothesis that worker behaviour is uncorrelated across successive brood care phases. Correlations were computed for all workers within colonies of the same size. The analysis could not be performed for colonies of size 2 because too few colonies of this size had a second brood-care phase.

Finally, we investigated whether behavioural fluctuations and task neglect decreased with group size. To compute behavioural fluctuations, we calculated colony mean r.m.s.d. by averaging the daily individual r.m.s.d. values of all colony members, computed the fluctuations (that is, absolute differences) of this colony mean r.m.s.d. between successive days, averaged these differences across the first brood-care phase and compared these mean fluctuations across colonies of different sizes.

Task neglect was computed using a r.m.s.d.-based proxy indicative of the consistent performance of tasks taking place in the nest. This metric, the minimum r.m.s.d., is the r.m.s.d. value of the ant with the highest spatial fidelity to the nest in each colony. The lower the minimum r.m.s.d., the more likely it is that at least one ant is at the nest—that is, that the brood is not left unattended. Task neglect was also quantified as the proportion of times in which the brood was unattended (that is, when no worker was found in the nest), using a combination of manual and automated tracking. To this aim, the position of the brood pile was annotated manually (if it could be determined by eye from images) every three days for each colony (Extended Data Fig. 9a). The distance between individual ant positions (obtained from automated tracking) and the brood pile was calculated for each frame in the previous three days (that is, distances were calculated between ant positions on days 1–3 and the brood-pile position on day 3). An ant was considered to be in the nest if it was within 5 mm of the brood-pile contour. For each colony, 'observed task neglect' was defined as the proportion of frames of the brood-care phase in which no ant was found in the nest (that is, task neglect implicitly refers to nursing here) (Extended Data Fig. 9b). Because larger colonies have higher ant density, the probability that at least one ant is found in the nest (as in any other area of the Petri dish) could increase with group size in a trivial way, without any associated change in individual behaviour. To control for this, each manually annotated brood area was also rotated by 180° around the centre of the Petri dish to produce a control area in the box of the same shape and area as the brood area (Extended Data Fig. 9a), and the number of ants in that random area was counted as above to produce a measure of 'expected task neglect' under the null model of no behavioural change (Extended Data Fig. 9b). If task neglect decreases with group size, we expected the difference between observed and expected task neglect (or 'effective task neglect') to decrease with group size (Extended Data Fig. 9c).

For all behavioural analyses, ants were excluded from the dataset if they were detected in less than 30% of the frames acquired within the considered time frame (brood-care phase or day; for ants that died during the brood-care phase, the considered time frame was the portion of the brood-care phase preceding death). This is unlikely to have introduced a bias because low-r.m.s.d. and high-r.m.s.d. workers have similar detection probabilities (see sample sizes in Extended Data Fig. 1f).

Statistical analyses. The effects of group size (1, 2, 4, 6, 12 or 16), genotype (A or B) and their interaction on behaviour (mean r.m.s.d., standard deviation of r.m.s.d., behavioural consistency, behavioural fluctuations, minimum r.m.s.d. and task neglect) and fitness (colony growth and time to eclosion) were investigated using generalized linear models. Colonies of size 1 were excluded from the models of standard deviation of r.m.s.d., behavioural consistency and minimum r.m.s.d. because the corresponding values were constant at 0, undetermined and uninformative, respectively. When needed, response variables were transformed to satisfy model assumptions of normally distributed residuals (tested with a Wilk–Shapiro test) and homoscedasticity (tested with Levene's test). We evaluated the significance of effects and their interaction by comparing pairs of nested models using χ^2 log-likelihood ratio tests following deletion of terms (starting with the interaction). Data from genotypes A and B were pooled whenever justified by the absence of a significant interaction term between the effect of genotype and the effect of group size. Statistical analyses were performed in R[32]. Full statistical results are presented as Supplementary Notes.

Theoretical model. First introduced to the social insect literature by Bonabeau et al.[23], fixed response thresholds have been a widely used approach to study the emergence of DOL in self-organized social systems. The model considers n individuals and assumes that there are two possible states for any given individual—active and inactive. Active individuals perform exactly one of m tasks at any moment in time (for simplicity, we assume that there are only two tasks, that is $m = 2$). Inactive individuals do not perform any task; they are considered to be in a rest state. An n by m binary matrix, $X_t = [x_{ij,t}]$, describes the activity and task state of each individual at a given time step t: if individual i is inactive, then all $x_{ij,t} = 0$ because an inactive individual performs no task; if individual i is active, then exactly one $x_{ij,t} = 1$ while all others are 0.

The model assumes that each task j has an associated stimulus, $s_{j,t}$, which signals the group-level demand for that task at time t. The change in the stimulus over discrete time can be modelled according to a previously published study[32] as:

$$s_{j,t+1} = s_{j,t} + \delta_j - \alpha \frac{\sum_{i=1}^{n} x_{ij,t}}{n}$$

in which δ_j is the constant, task-specific stimulus increase rate per time step (that is, task demand rate), α is a scalar measuring task performance efficiency (assumed, for simplicity, to be the same for all individuals across all tasks), $\sum_{i=1}^{n} x_{ij,t}$ is the number of individuals performing task j at time t and n is the total number of individuals in the colony. The higher the α, the better the individuals are at performing the tasks. The higher the δ_j, the more demanding that task; for simplicity, all tasks are assumed to have the same demand rate, δ.

The model has four sources of stochasticity. First, each individual i has an internal threshold, θ_{ij}, for each task j. This is randomly drawn from a normal distribution with mean μ_j and normalized standard deviation σ_j, which is given as a proportion of μ_j (for example, $\sigma = 0.3$ indicates a standard deviation that is 30% of the mean). For simplicity, we assume that μ and σ are the same for all tasks. Second, inactive individuals are exposed to task stimuli randomly in a given time step, until they either commit to performing a given task and thus become active, or cycle through all stimuli without becoming active and thus remain inactive. Third, for each encountered stimulus, individuals determine whether to perform that task by evaluating the stimulus level relative to their corresponding internal threshold. The threshold response function that gives the probability $P_{ij,t}$ that individual i performs task j at time t is a sigmoid for which the steepness is determined by a parameter η to range from more deterministic to more stochastic[33]:

$$P_{ij,t} = \frac{s_{j,t}^{\eta}}{s_{j,t}^{\eta} + \theta_{ij}^{\eta}}$$

For large values of η, we recover a deterministic behaviour, such that a task is only performed if the stimulus exceeds the threshold, and in that case it is always performed. In our simulations, we vary $1 \leq \eta \leq 30$ to capture this range of behaviours. Fourth, upon starting a task, an individual will continue performing that task until it spontaneously quits, with a constant quit probability, τ[9,10,23,34]. Active individuals do not evaluate task stimuli and do not switch between tasks; only inactive individuals evaluate stimuli and determine which task (if any) to start performing.

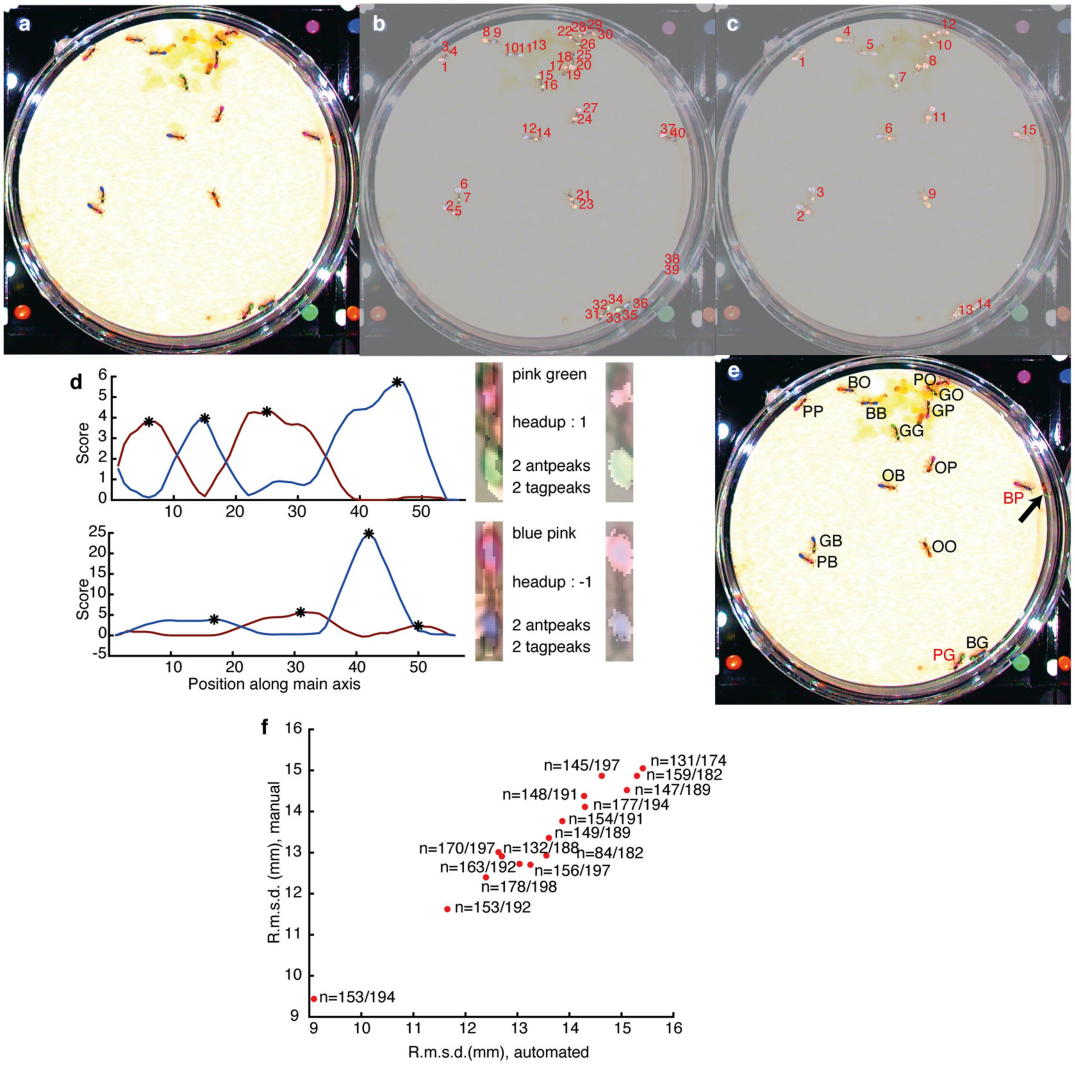
To analyse this model and—specifically—how each of the four sources of stochasticity affects the outcome, we started from a fully deterministic version and built in each one of the different sources of stochasticity independently (Extended Data Fig. 7, Supplementary Methods). All agent-based simulations and subsequent data analyses were conducted in R[32].

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

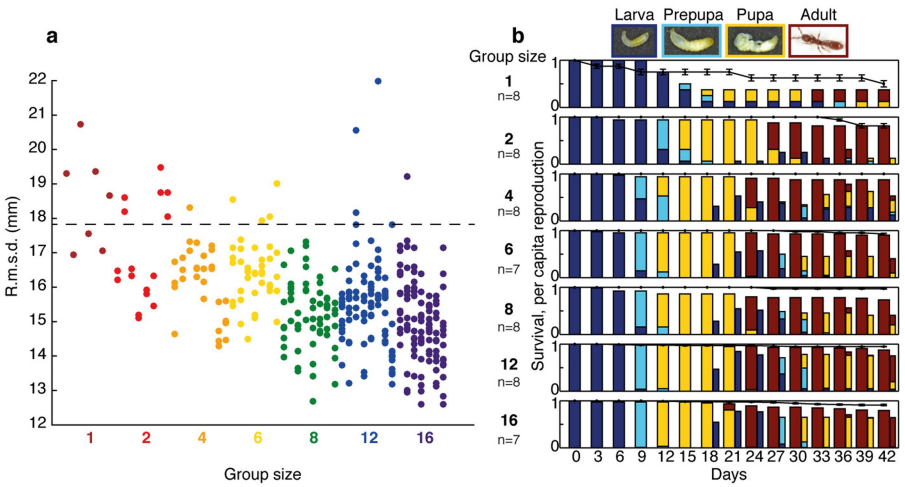
Code availability. All behavioural tracking code is available at https://doi.org/10.5281/zenodo.1211644. All code for model simulations is available at https://doi.org/10.5281/zenodo.1211231.

Data availability. All behavioural tracking data (x, y positions of individual ants in different frames) as well as colony summary statistics (behaviour and fitness) are available at https://doi.org/10.5281/zenodo.1237867. Any other data that support the findings of this study, such as processed data files used for statistical analyses, are available from the corresponding authors upon reasonable request.

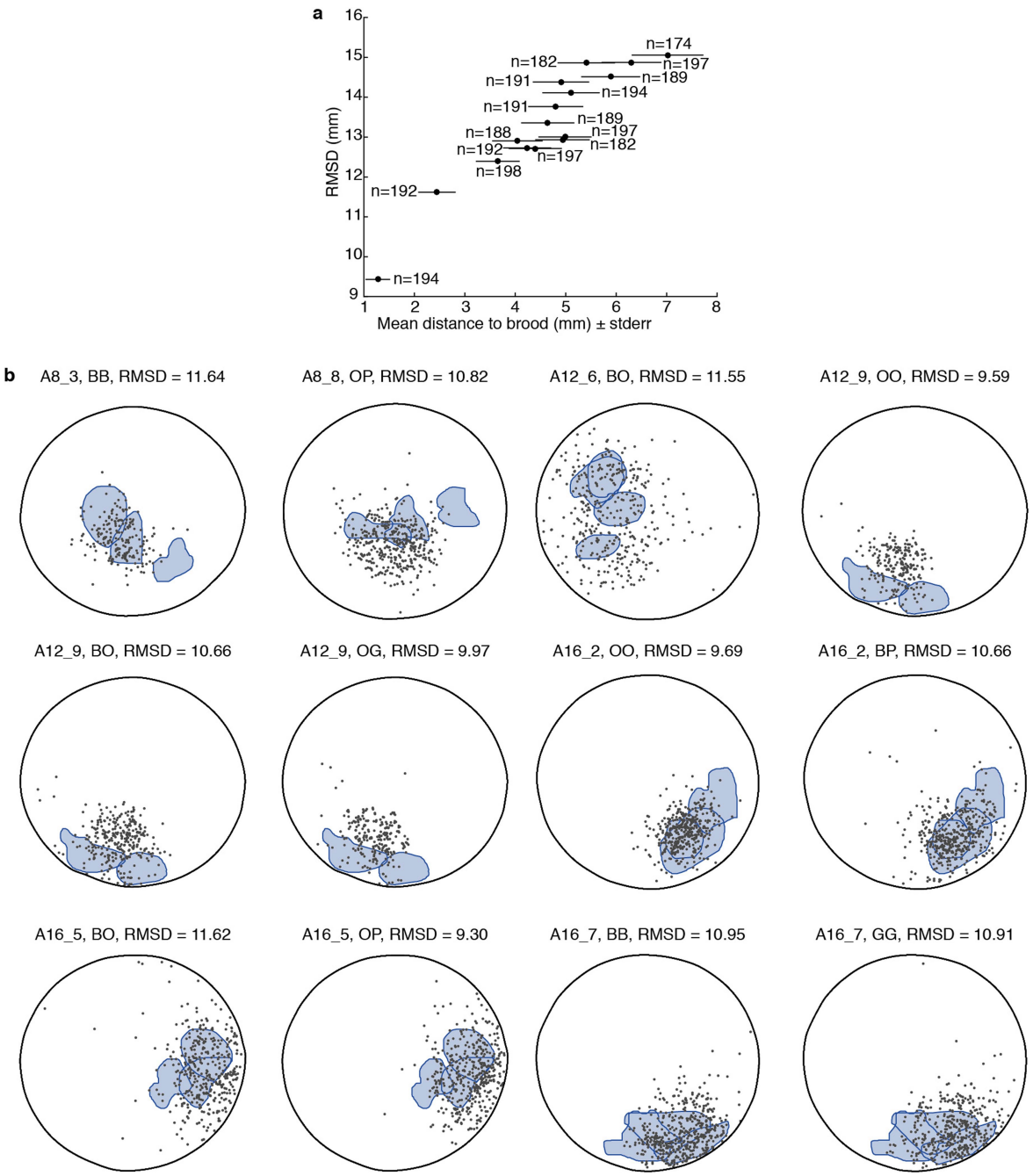
31. Ravary, F. & Jaisson, P. The reproductive cycle of thelytokous colonies of *Cerapachys biroi* Forel (Formicidae, Cerapachyinae). *Insectes Soc.* **49**, 114–119 (2002).
32. R Core Team. *R: A Language and Environment for Statistical Computing* http://www.R-project.org/ (R Foundation for Statistical Computing, Vienna, 2008).
33. Dodds, P. S. & Watts, D. J. Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.* **92**, 218701 (2004).
34. Bonabeau, E., Theraulaz, G. & Deneubourg, J.-L. Fixed response thresholds and the regulation of division of labor in insect societies. *Bull. Math. Biol.* **60**, 753–807 (1998).



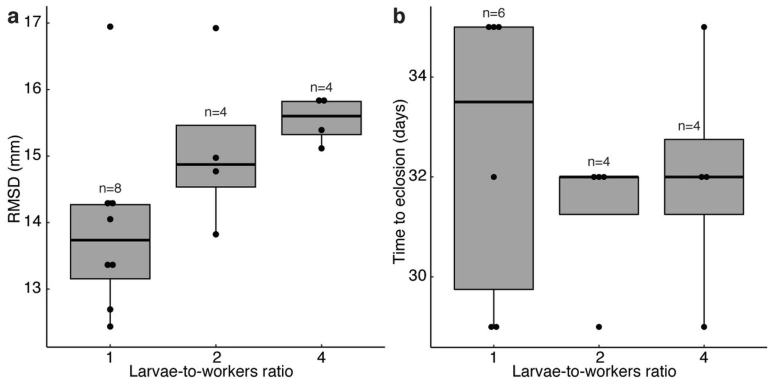
Extended Data Fig. 1 | Ant detection algorithm. a, Example cropped frame showing one 16-worker colony after image correction. **b**, Colour-tag detection. The highlighted numbered zones are image regions containing pixels that were assigned a high probability for tag colours (green, blue, orange or pink) by a Bayesian classifier. **c**, Candidate ant detection. The highlighted numbered zones correspond to contiguous regions containing colour tags and pixels that were assigned a high probability for ant colour (that is, cuticle) by the classifier. **d**, Candidate ant orientation. Candidate ants are aligned using the segment connecting the two colour tags, and oriented (head down versus head up) on the basis of the relative position of cuticle- and tag-probability maxima (black stars on brown and blue lines, respectively) along the main axis of each region. **e**, Final IDs after using Munkres' variant of the Hungarian assignment algorithm. Labels indicate colour IDs (thorax–abdomen; G, green; B, blue; O, orange; P, pink). Ants shown as examples in **d** are labelled in red. All assignments shown are correct, but one ant is missed (arrow). This panel is identical to Fig. 1a. **f**, Correlation between r.m.s.d. calculated from automated versus manual assignments for one 16-worker colony. The r.m.s.d. was computed from a subset of frames in the brood-care phase. $n =$ (number of automated detections)/(number of manual detections). Pearson's $r = 0.95$, $P < 0.001$.



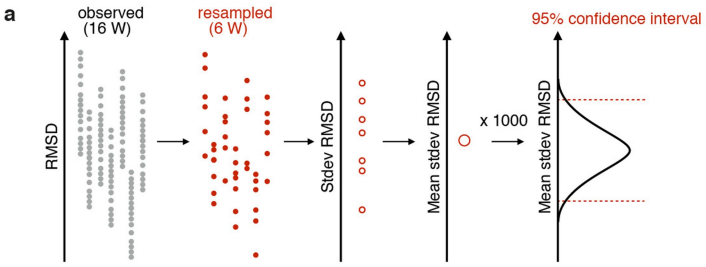
Extended Data Fig. 2 | The r.m.s.d. and fitness in genotype B. **a**, Individual r.m.s.d. values for all workers of genotype B. Ants from the same colony are vertically aligned. The dashed line represents the expected r.m.s.d. assuming a uniform distribution of an ant's positions. **b**, The dynamics of brood development as a function of group size in genotype B. The proportion of the brood in successive developmental stages (colours) in colonies of sizes 1–16 is shown. Wide and narrow bars indicate first and second brood generations, respectively. Black line, worker survival (mean ± s.e.m.).



Extended Data Fig. 3 | The r.m.s.d. and spatial fidelity to the nest. **a**, Correlation between individual r.m.s.d. and individual distance to the brood (mean ± s.e.m.) over the first brood-care phase in one colony of 16 workers (Spearman's $r = 0.93$, $P = 0$; $n = 16$ ants). Behavioural traits are based on 209 manually tracked frames. Sample sizes (n) indicate the number of frames in which each ant was manually identified. Manual tracking was used here because the automated tracking algorithm does not allow us to locate the brood. **b**, Individuals with low r.m.s.d. (r.m.s.d. < 12 in Fig. 1b) have high spatial fidelity to the nest area. Each circle represents the spatial distribution of an ant (grey dots) with respect to the brood pile (shaded blue areas) in the brood-care phase. Panel titles indicate colony identity (for example, A8_3 is the third replicate colony of genotype A and size 8), ant identity (for example, BO for blue–orange) and individual r.m.s.d. In each colony, the brood pile was manually annotated every three days (that is, if the brood-care phase lasted nine days, three brood piles zones were manually annotated; zones could overlap or not, depending on how much the brood pile moved).



Extended Data Fig. 4 | Effect of the larvae-to-workers ratio on behaviour and brood developmental time. The number of workers was constant at 4, and the number of larvae varied between 4 and 16, so as to obtain larvae-to-workers ratios of 1, 2 or 4. **a**, Mean colony r.m.s.d. increased with the larvae-to-workers ratio (log-transformed r.m.s.d.: $\chi^2 = 5.00$, $P = 0.03$). **b**, Larval time to eclosion was unaffected by the larvae-to-workers ratio (time to eclosion transformed by (time to eclosion)^5: $\chi^2 = 0.17$, $P = 0.68$). Sample sizes indicate the number of colonies in which at least one larva reached adulthood. In both panels, box plots represent the median (thick horizontal line); the lower and upper hinges correspond to the first and third quartile, respectively. The upper whiskers extend from the upper hinge to the largest value no further than 1.5× interquartile range from the hinge; the lower whiskers extend from the lower hinge to the smallest value no further than 1.5× interquartile range from the hinge.



b

phase:	brood care 1				reproductive 1			brood care 2						reproductive 2			
day:	**1**	**2**	**…**	**10**	**11**	**…**	**17**	**18**	**19**	**…**	**…**	**30**	**31**	**32**	**…**	**39**	**40**
	20.8	20.4	…	4.6				4.9	4.7	…	…	6.9	3.7				
	20.9	19.5	…	5.7				6.0	5.2	…	…	9.4	5.3				
…	…	…	…	…				…	…	…	…	…	…				
	19.5	16.9	…	4.4				5.7	2.2	…	…	5.7	5.8				

c

phase:	brood care 1				reproductive 1			brood care 2						reproductive 2			
day:	**1**	**2**	**…**	**10**	**11**	**…**	**17**	**18**	**19**	**…**	**…**	**30**	**31**	**32**	**…**	**39**	**40**
	6	7	…	3				1	3	…	…	4	2				
	7	6	…	7				5	5	…	…	5	4				
…	…	…	…	…				…	…	…	…	…	…				
	1	1	…	2				3	1	…	…	1	5				

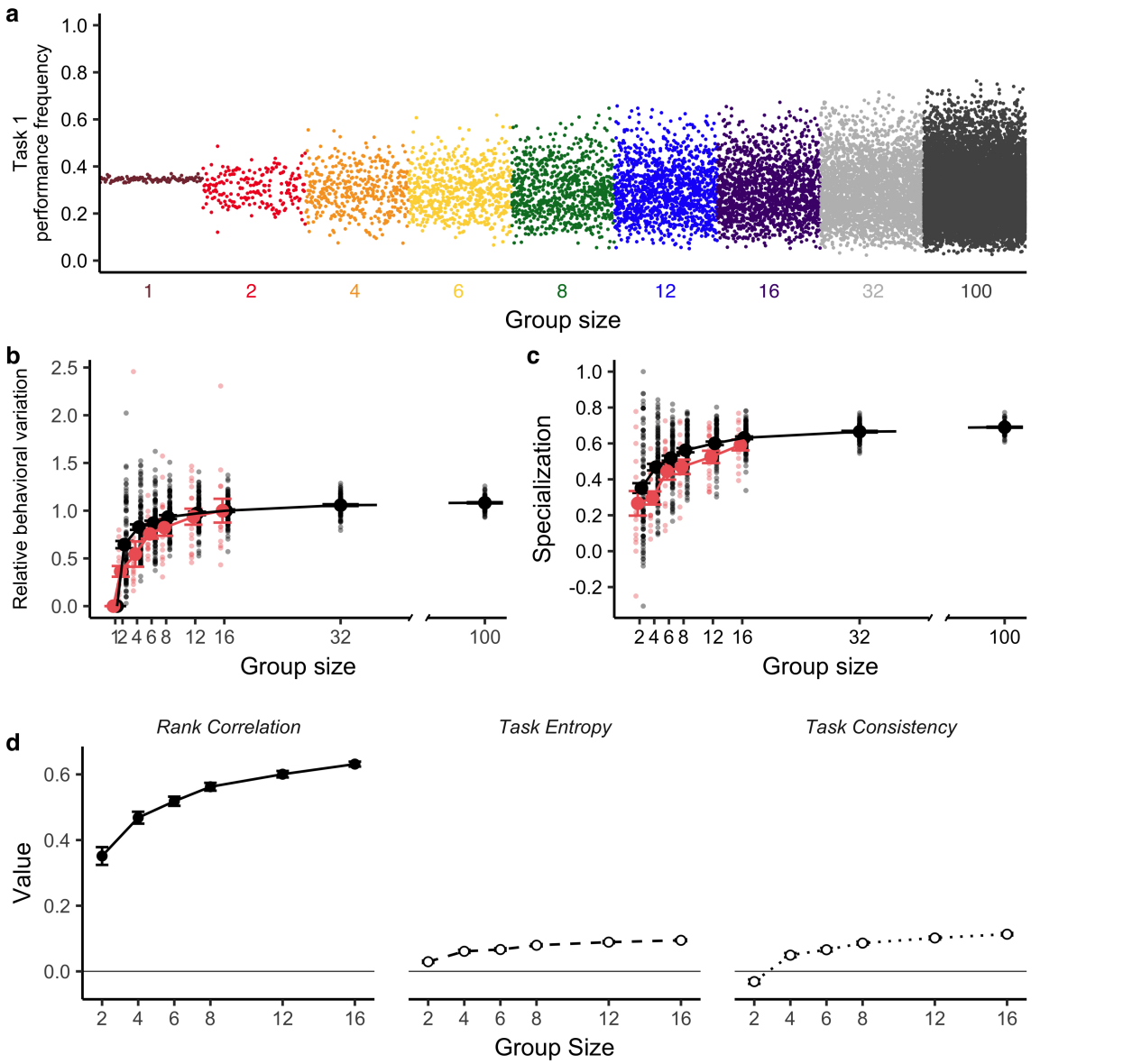
d

phase:	brood care 1				
day:	**1**	**2**	**…**	**9**	**10**
1	1		…		
2	0.9	1	…		
…	…	…	1		…
9	0.8	0.4	…	1	
10	0.3	0.4	…	-0.1	1

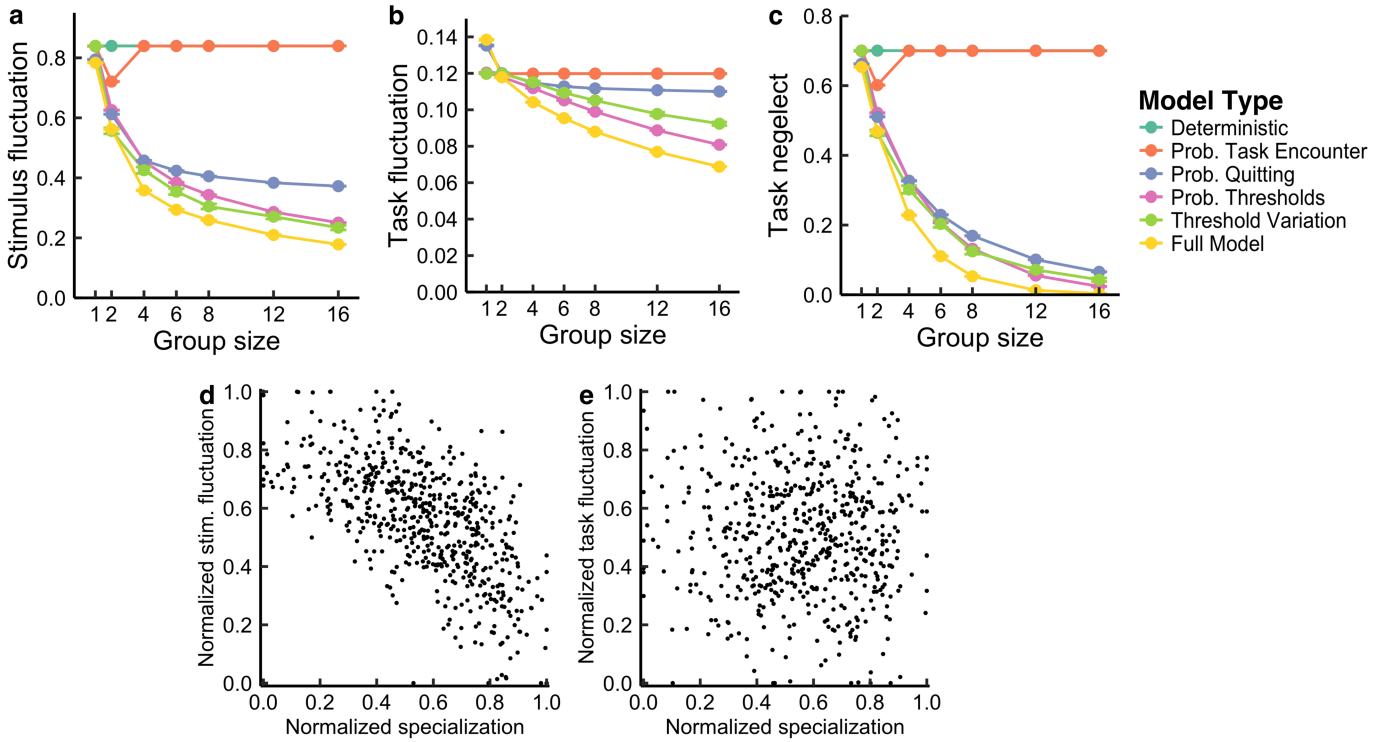
e

phase:	brood care 1			reproductive 1			brood care 2			reproductive 2		
day:	**1**	**…**	**10**	**11**	**…**	**17**	**18**	**…**	**31**	**32**	**…**	**40**
	5						7					
	4						6					
…	…						…					
	6						1					

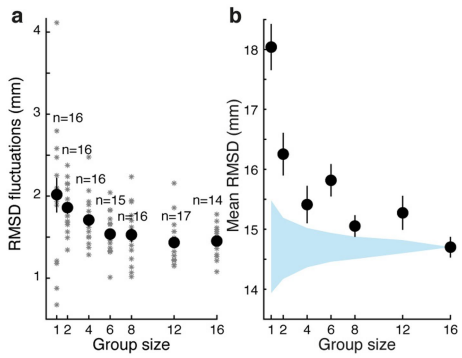
Extended Data Fig. 5 | Methods for behavioural analyses. a, Resampling scheme. Ninety-five per cent confidence intervals were generated by resampling individual r.m.s.d. values from one colony of size 16 at a time. In the example here, the generation of confidence intervals for behavioural variation (standard deviation of r.m.s.d.) in colonies of size 6 is shown. The same method was used to generate confidence intervals for mean colony behaviour (mean r.m.s.d.). **b**–**e**, Computing specialization. **b**, Daily individual r.m.s.d. values in one colony of size 8. **c**, Daily individual r.m.s.d. ranks. **d**, Pairwise rank correlation matrix between days of the first brood-care phase. Values highlighted in red indicate rank correlations (Spearman, $n = 16$ ants) between consecutive days, which are averaged to compute short-term behavioural consistency. **e**, Mean r.m.s.d. ranks per brood-care phase used to compute long-term behavioural consistency.



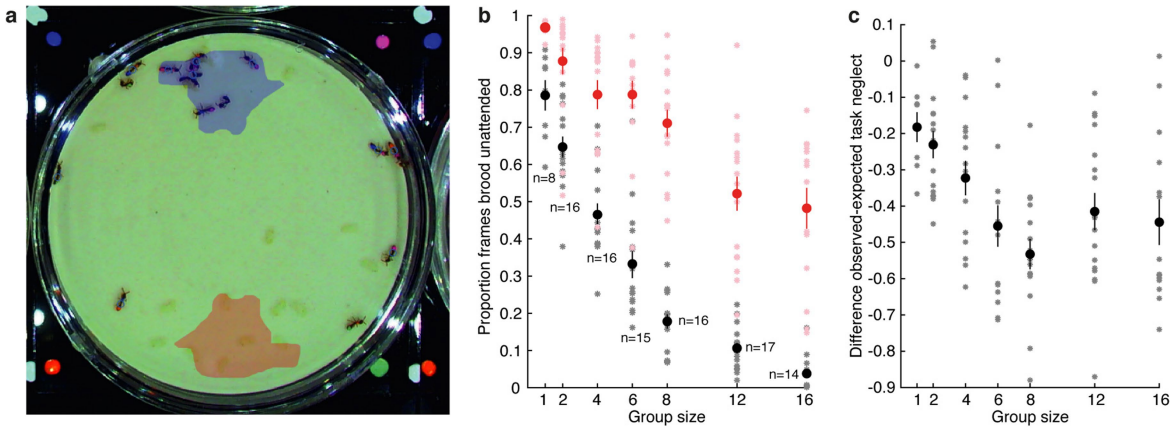
Extended Data Fig. 6 | Behaviour of the fixed threshold model. One hundred replicates were simulated per group size. Parameterization: $m = 2$, $\eta = 7$, $\mu = 10$, $\sigma = 0.1$, $\tau = 0.2$ and $\delta = 0.6$, corresponding to the filled circle symbol in Fig. 3b. **a**, Frequency of task 1 performance (measured across a simulation run) by individual ants at different group sizes; each point represents an ant and ants from the same colony are vertically aligned. **b**, Behavioural variation (standard deviation of individual task performance frequencies) across all 100 replicates for each group size, averaged over both tasks and shown relative to group size 16. **c**, Specialization in task performance relative to group size. Each point represents one colony, and the line represents the mean value (±s.e.m.) across all 100 replicates for each group size. In **b**, **c**, model output is in black and experimental data are in red. **d**. Mean values (±s.e.m.) of the rank correlation, task entropy and task consistency metrics across all 100 replicates at each group size.



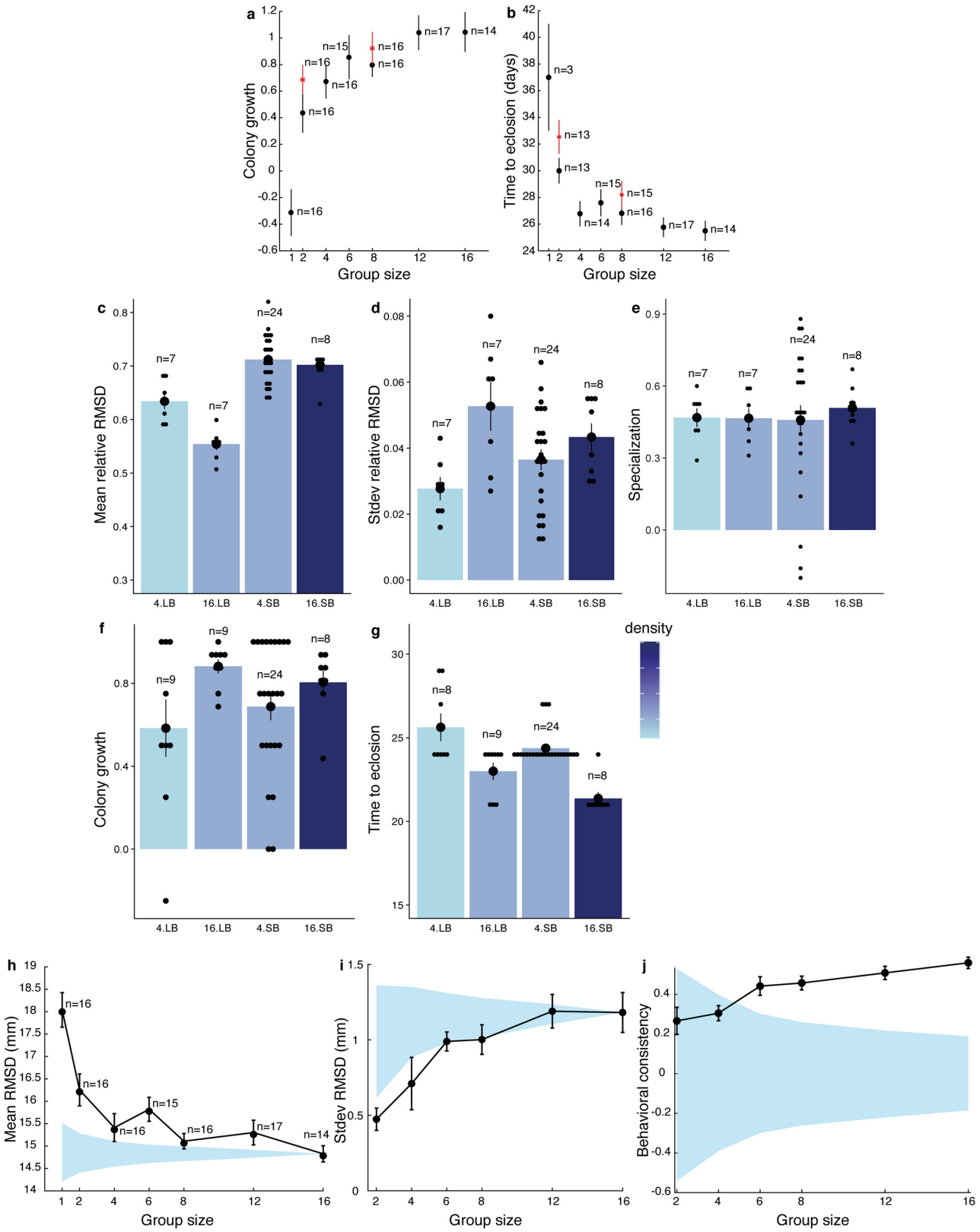
Extended Data Fig. 7 | The effect of stochasticity and specialization on proxies for fitness. One hundred replicates were simulated per group size. Parameter settings for the deterministic model can be found in the Supplementary Methods; departures from deterministic model parameters are as in Extended Data Fig. 6. **a**, Short-term (single time step) stimulus fluctuations averaged across both tasks are shown across group sizes and for all models. **b**, Short-term (single time step) fluctuations in task performance frequency (measured by the proportion of the colony performing each task), averaged across both tasks, are shown across group sizes and for all models. **c**, Task neglect averaged across both tasks is shown across group sizes and for all models. In **a**–**c**, points represent the described averages, which have been further averaged (mean ± s.e.m.) across $n = 100$ replicate colonies of a given size. **d**, **e**, Relationship between specialization and short-term stimulus fluctuations (**d**) or short-term fluctuations in task performance frequency (**e**), in the full model when controlling for group size. Each point represents one simulated colony.



Extended Data Fig. 8 | Behavioural homeostasis increases with group size. a, Day-to-day fluctuations in colony mean r.m.s.d. (mean ± s.e.m.) decrease with group size ($\chi^2 = 21.30$, $P = 3.93 \times 10^{-6}$). Asterisks represent colony-level data. **b**, Mean spatial fidelity increases with group size. Black, colony mean r.m.s.d. as a function of group size (mean ± s.e.m.). Blue, 95% confidence intervals under the null hypothesis of no group-size effect on individual behaviour, generated by resampling individuals from colonies of size 16 (Extended Data Fig. 5a). Sample sizes are as in **a**. In both panels, data for genotypes A and B are pooled.



Extended Data Fig. 9 | Task neglect. a, Manually annotated nest area (blue) and control area (red) generated by rotating the nest area by 180° around the centre of the Petri dish. **b**, Task neglect (mean ± s.e.m.) decreases with group size. The proportion of frames in which no ant was found near the brood as a function of group size. Black, observed task neglect. Red, expected task neglect. **c**, Effective task neglect (mean ± s.e.m.) decreases with group size ($\chi^2 = 13.36$, $P = 2.57 \times 10^{-4}$). The difference between observed and expected task neglect is shown as a function of group size. Sample sizes are as in **b**. In **b**, **c**, data for genotypes A and B are pooled and asterisks represent colonies.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Control experiments. a, **b**, Paint-marking did not disproportionately affect small colonies. Red asterisks indicate control colonies composed of unmarked ants; otherwise, data are as in Fig. 4b, c. **a**, Growth in colonies of unmarked ants (mean ± s.e.m.). Colony growth was unaffected by paint-marking ($\chi^2 = 2.71$, $P = 0.10$), the interaction of paint-marking with group size ($\chi^2 = 0.31$, $P = 0.58$) or the interaction of paint-marking with genotype ($\chi^2 = 0.17$, $P = 0.68$). **b**, Larval time-to-eclosion in colonies of unmarked ants (mean ± s.e.m.). Time to eclosion of larvae was increased by paint-marking of the workers (square-root-transformed time to eclosion: $\chi^2 = 8.98$, $P = 0.003$), but paint-marking did not interact with group size ($\chi^2 = 0.09$, $P = 0.77$) or genotype ($\chi^2 = 0.22$, $P = 0.64$). **c**–**g**, Effects of density on behaviour and fitness. Colonies consisted of 4 or 16 workers (and a matching number of larvae) in small or large Petri dishes (SB and LB, respectively), corresponding to 3 densities (shades of blue). **c**, Mean spatial fidelity (mean ± s.e.m.) was affected by group size ($\chi^2 = 6.49$, $P = 0.01$), box size ($\chi^2 = 38.46$, $P = 5.6 \times 10^{-10}$) and density (group size:box size: $\chi^2 = 6.76$, $P = 0.009$). **d**, Behavioural variation (mean ± s.e.m.) was affected by group size ($\chi^2 = 7.44$, $P = 0.006$) but not by box size ($\chi^2 = 0.08$, $P = 0.77$) or density (group size:box size: $\chi^2 = 3.50$, $P = 0.06$). **e**, Behavioural consistency (mean ± s.e.m.) was not affected by group size ($\chi^2 = 0.03$, $P = 0.87$), box size ($\chi^2 = 0.22$, $P = 0.64$) or density (group size:box size: $\chi^2 = 0.02$ $P = 0.88$). Behavioural consistency was transformed by (behavioural consistency + 0.21)$^{1.5}$. **f**, Colony growth (mean ± s.e.m.) was affected by group size ($\chi^2 = 3.91$, $P = 0.048$), but not by box size ($\chi^2 = 0.04$, $P = 0.85$) or density (group size:box size: $\chi^2 = 1.00$, $P = 0.32$). Colony growth was transformed by (growth + 0.4)$^{1.9}$. Thus, the effect of density is small relative to that of group size, and variation in density alone is therefore very unlikely to have confounded our results. **g**, Larval time-to-eclosion (mean ± s.e.m.) was affected by group size ($\chi^2 = 35.74$, $P = 2.26 \times 10^{-9}$) and box size ($\chi^2 = 10.45$, $P = 0.001$) but not by density (group size:box size: $\chi^2 = 0.67$, $P = 0.41$). Time to eclosion was transformed by (time to eclosion)$^{-0.3}$. **h**–**j**, Removing individuals with more than three ovarioles from analyses did not qualitatively affect our results. **h**, Mean spatial fidelity of the colony increases with group size. Black, mean r.m.s.d. (±s.e.m.) as a function of group size, after excluding individuals with four or more ovarioles. Blue, 95% confidence interval generated by resampling workers from 16-worker colonies (Extended Data Fig. 5a). **i**, Behavioural variation increases with group size. Black, standard deviation in r.m.s.d. per colony as a function of group size (mean ± s.e.m.), after excluding individuals with more than three ovarioles. Ninety-five per cent confidence intervals and sample sizes are as in **a**. **j**, Day-to-day rank consistency increases with group size. Black, mean r.m.s.d. rank correlation coefficients over consecutive days in the first brood care phase as a function of group size (mean ± s.e.m.), after excluding individuals with more than three ovarioles. Blue, 95% confidence intervals generated by randomizing daily ranks in each colony. In **a**, **b**, **h**–**j**, data for genotypes A and B are pooled.

Corresponding author(s): Daniel Kronauer (dkronauer@rockefeller.edu), Yuko Ulrich (yuko.ulrich@gmail.com)

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	All behavioral tracking code is available at http://doi.org/10.5281/zenodo.1211644. All code for model simulations is available at http://doi.org/10.5281/zenodo.1211231.
Data analysis	*Computational modelling was performed in R v.3.5.0. Empirical data analyses were performed in MATLAB v.2017b and R v.3.4.3.*

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All behavioral tracking data (x, y positions of individual ants in different frames) as well as colony summary statistics (behavior and fitness) are available at https://doi.org/10.5281/zenodo.1237867. Any other data that support the findings of this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We used the maximum possible sample size (i.e. number of replicate colonies), which was ultimately only constrained by the maximum size and processing capacity of our custom tracking setup.
Data exclusions	Two colonies (size 6 and 16, genotype B) were excluded from all analyses due to setup errors (incorrect number of workers or larvae at the beginning of the experiment). For all behavioral analyses, individual ants were excluded from the dataset if they were detected in less than 30% of the frames acquired within the considered time frame.
Replication	The experiment was replicated in two independent ant genotypes.
Randomization	Individual ants were collected from a stock colony, assigned color-tags and distributed across experimental colonies at random. The order in which the colonies were set up, as well as their position in the tracking setup, were randomized.
Blinding	Investigators did not directly participate in data acquisition and analysis (behavioral tracking), because these processes were automated.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging