HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2019 February 22.

Published in final edited form as:

Nature. 2018 August ; 560(7720): 635–638. doi:10.1038/s41586-018-0422-6.

"Fitness benefits and emergent division of labor at the onset of group-living"

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

[1]Laboratory of Social Evolution and Behavior, The Rockefeller University, New York, NY 10065, USA. [2]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland. [3]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA.

Abstract

The initial fitness benefits of group-living are considered the greatest hurdle to the evolution of sociality[1], and theory predicts that they 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, they have been proposed to stem from division of labor (DOL)[5–8], which is characterized by between-individual variability and within-individual consistency (specialization) in task performance. At the onset of sociality, however, groups were likely small and composed of similar individuals with potentially redundant rather than complementary function[1]. Theory suggests that DOL can emerge even in relatively small, simple groups[9,10]. However, empirical data on the effects of group size on DOL and fitness remain equivocal[6]. Here, we use long-term automated behavioral tracking in clonal ant colonies, combined with mathematical modeling, to show that increases in social-group size can generate DOL among extremely similar workers, in groups as small as six individuals. These early effects on behavior 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 higher DOL. Overall, our results indicate that DOL, increased homeostasis, and higher fitness can naturally emerge in small,

Correspondence and requests for materials should be addressed to Y.U. (yuko.ulrich@gmail.com) and D.J.C.K. (dkronauer@rockefeller.edu).

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 analyzed by Y.U. Theoretical modeling was performed by C.K.T. and C.E.T. Computational modeling 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.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

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 http://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.

Code availability

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.

The authors declare no competing financial interests.

homogeneous social groups, and that scaling effects associated with increasing group size can thus promote social cohesion at incipient stages of group-living.

Quantifying the effects of group size on DOL and fitness requires the ability to precisely manipulate group size, monitor individual behavior within groups, and accurately measure fitness in controlled conditions. Crucially, group size must be controlled independently from other factors that often co-vary with group size and can affect DOL and fitness, such as colony genetic or age structure[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, but 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 behavior and fitness of colonies containing 1 to 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 behavior 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[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 analyzed behavior 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] (e.g. foraging and waste disposal occur away from the nest, while nursing occurs at the nest), individual behavior can be described in terms of spatial location[18]. This is commonly done by assigning individuals to discrete behavioral groups based on manually acquired spatial data[13]. While the acquisition of spatial data has greatly improved with automated behavioral tracking[18–20], individuals are often still clustered into discrete behavioral groups[18]. However, in many systems—especially those without morphological castes—individual behavior is continuously distributed[21,22]. We therefore analyzed behavior non-parametrically from continuous spatial data, avoiding assumptions about the statistical distribution of individual behavior. The spatial distribution of each ant was measured as the 2-dimensional root-mean-square deviation (RMSD) of its x,y coordinates, i.e. the spread of those coordinates around their center of mass, throughout the brood care phase (Fig. 1b, Extended Data Fig. 2a; Methods). An ant’s RMSD captures its tendency to explore the arena, i.e. its lack of spatial fidelity (Fig. 1c), and strongly correlates with its mean distance to the nest (Extended Data Fig. 3).

RMSD is therefore a biologically-meaningful metric that reflects the propensity to perform tasks away from the nest (e.g. foraging) rather than at the nest (e.g. nursing). In fact, a colony's mean RMSD 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 RMSD, but did not cluster into discrete behavioral groups (Fig. 1b).

If DOL increases with group size, larger colonies are expected to show 1) higher behavioral variation between colony members and 2) higher individual behavioral consistency, or specialization, over time. Although not always independent from each other, these measures reflect distinct facets of DOL. Behavioral variation, computed as the standard deviation across RMSD values of ants from the same colony, indeed increased with group size (Fig. 1d, Extended Data Fig. 5a), with small colonies (sizes 2–8) displaying less behavioral variation than larger colonies (sizes 12–16). Short-term specialization was quantified as the RMSD rank correlation between consecutive days, averaged over the first brood care phase (Fig. 2a, Extended Data Fig. 5b-d); this captures the day-to-day behavioral 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-RMSD ranks in the first and second brood care phases (Extended Data Fig. 5e), was found in colonies of 6 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 modeling 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 reflecting colony demand based on innate, fixed thresholds that determine their propensity 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 behavioral continuum akin 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, including regions with a close quantitative match with our empirical observations (Figs. 3a-b; Supplementary Methods). Behavioral 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—e.g., 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 (Extended Data Fig. 6d), but the rank correlation metric (Fig. 2b) produced the highest values[10,26], suggesting it might be more sensitive to rudimentary types of specialization. Other metrics, however, also reveal important insights: e.g., task consistency, which quantifies how infrequently individuals switch between tasks (Supplementary Methods), increased with group size, even in the absence of task-switching costs, suggesting that reduced task switching could be an early emergent property.

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, i.e. instances when tasks are not performed by any ant (Fig. 3d). These effects were obtained by increasing group size alone, even in the absence of DOL, i.e. when there was no threshold variation, as long as some other source of stochasticity reduced the likelihood of behavioral 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), though not all of these effects (Extended Data Fig. 7e). 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 behavior (Extended Data Fig. 8a), increases in colony-level spatial stability (Extended Data Fig. 8b), and decreases in minimum RMSD (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 (e.g. larval hunger) and more consistent work performance (e.g. fewer instances of unattended brood). Since colony homeostasis is considered 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 was negative for the smallest colonies but rapidly increased with group size and plateaued at around 1 in colonies of sizes 12 and 16 (Fig. 4b), similar to values reported in colonies orders-of-magnitude larger[27]. Differences in colony growth were partially due to unexpected effects on brood development: time to eclosion was 10 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 likely 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 imply that small colonies had prolonged cycles. In fact, some large colonies had produced two cohorts of workers by the time small colonies 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 emerge as a function of group size at incipient stages of group-living. Importantly, the rudimentary, flexible DOL demonstrated here does not rely on well-known mechanisms like morphological caste specialization or age polyethism[12], but instead on plastic behavioral responses to the social environment[28–30]. While our theoretical model shows that specialization alone can increase group 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, like those with morphologically differentiated queen and worker castes.

Methods

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-days old larvae in airtight Petri dishes (5cm in diameter, corresponding to ca. 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 harvested from the same cohort and had eclosed within a day of each other (due to the synchronized reproduction of *O. biroi*). From the time they were harvested (1–3 days post-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 adults. However, we cannot exclude that small differences in individual experience occurred even in this common environment before the start of the experiment. All workers were tagged with color 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 (i.e., large, healthy) laboratory stock colony in the brood care phase. The experiment was conducted in two distinct genotypes, A and B[15]. 7–9 replicate colonies were used for each group size and genotype, for 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 behavior is not affected by light). Every 3 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 (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). 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

Behavioral 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 4 colonies, and the position of colonies within the setup was randomized. This resulted in 7976 and 6429 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 PCs used to control the webcams (median interval between frames: 420 seconds for genotype A, 525 seconds for genotype B), and in part from a ca. 35 hour long scanning interruption in the genotype B experiment. This interruption occurred in the reproductive phase of most colonies (51 out of 56) colonies. Because behavioral analyses were conducted on data collected in the brood care phase only, this interruption did not affect our results and conclusions. Fitness monitoring outlasted behavioral 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 8 following color categories: plaster, ant cuticle, shadow, food, and color tags (pink, orange, blue, and green) (Extended Data Fig. 1b). Size and color-probability thresholds were used to detect candidate regions corresponding to ants carrying color-tags (Extended Data Fig. 1c). Candidate ants were oriented based on the relative position of cuticle and tag color-probability maxima along the main axis of each region (e.g. 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 color 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 (i.e. 22.9% of ants were missed). Of all the automated assignments, 94.4% were correct (i.e. 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 behavioral traits computed from automated assignments correlate with the same traits computed from manual assignments.

Individual behavioral RMSD 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®.

Behavioral data analysis

We restricted our behavioral analyses to the brood care phase because worker locomotion, and thus our ability to detect inter-individual behavioral differences, is dramatically 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 (i.e. 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 Ravary and Jaisson[31], who characterized the brood care phase by 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 2-dimensional root-mean square deviation (RMSD):

$$RMSD = \sqrt{\frac{\Sigma_i\left(\left(x_i - \bar{x}\right)^2 + \left(y_i - \bar{y}\right)^2\right.}{n}}$$

where (x_i,y_i) are the coordinates of the focal ant in frame i, $(\bar{x},\ \bar{y})$ are the coordinates of the center 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. RMSD 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 (e.g. nursing the larvae) and little time performing outside tasks (foraging, waste disposal) have low RMSD values, while workers that spend comparatively more time away from the brood have higher RMSD values. In previous studies, workers displaying behavior corresponding to low or high RMSD have been labeled 'nurses' and 'foragers', respectively. However, given the apparent continuous distribution of RMSD values across individuals in this study (Fig. 1b), we chose not to cluster individuals into discrete behavioral 'castes'.

For each colony, mean behavior was computed as the average of individual RMSD values, and behavioral variability was computed as the standard deviation of individual RMSD 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 where 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 behavior 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 behavior and behavioral variability were calculated for each simulated colony and averaged across replicate colonies of a given size, as described above. This resampling procedure was repeated 1000 times. 95% confidence intervals were generated for mean behavior and behavioral variation for each group size separately using

the same resampled data, to test which colony size had an observed behavior 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 behavioral 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 behavioral data (RMSD): colony-level behavioral consistency, or specialization, was defined for each colony as the correlation coefficient between individual RMSD-ranks on consecutive days, averaged over the first brood care phase (Extended Data Fig. 5b-d). Spearman rank correlations, rather than parametric correlations (e.g. Pearson) were used because there was more variation in RMSD over time than across individuals (Fig. 2a). The timeframe of days was chosen to ensure that individual RMSD 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 1000 times, based on the null hypothesis that worker behavior is uncorrelated across successive days. We then tested whether individual behavior 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 4 days of behavioral data were available. In these colonies, for each brood care phase, workers were assigned a within-colony rank based on their mean RMSD across days (Extended Data Fig. 5e). Long-term behavioral 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 1000 times, based on the null hypothesis that worker behavior is uncorrelated across successive brood care phases. Additionally, 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 behavioral fluctuations and task neglect decreased with group size. To compute behavioral fluctuations, we calculated colony mean RMSD by averaging the daily individual RMSD values of all colony members, computed the fluctuations (i.e. absolute differences) of this colony mean RMSD 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 RMSD-based proxy indicative of the consistent performance of tasks taking place in the nest. This metric, the minimum RMSD, is the RMSD value of the ant with the highest spatial fidelity to the nest in each colony. The lower the minimum RMSD, the more likely it is that at least one ant is at the nest, i.e. that the brood is not left unattended. Task neglect was also quantified as the proportion of times

where the brood was unattended (i.e. 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 3 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 3 days (i.e. 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 5mm 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 (i.e. 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 behavior. To control for this, each manually annotated brood area was also rotated by 180° around the center 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 behavioral 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 increase with group size (Extended Data Fig. 9c).

For all behavioral 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-RMSD and high-RMSD 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 behavior (mean RMSD, stdev RMSD, behavioral consistency, behavioral fluctuations, minimum RMSD, task neglect) and fitness (colony growth, time to eclosion) were investigated using generalized linear models (GLM). Colonies of size 1 were excluded from the models of stdev RMSD, behavioral consistency, and minimum RMSD 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 Chi-square log-likelihood ratio tests (LRTs) 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. 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 $x_{ij,t}=0$ since 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 modeled according to Bonabeau *et al.*[32] as:

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

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

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 (e.g., σ=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 thereby become active, or cycle through all stimuli without becoming active and thereby 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 whose steepness is determined by a parameter η to range from more deterministic to more stochastic[33]:

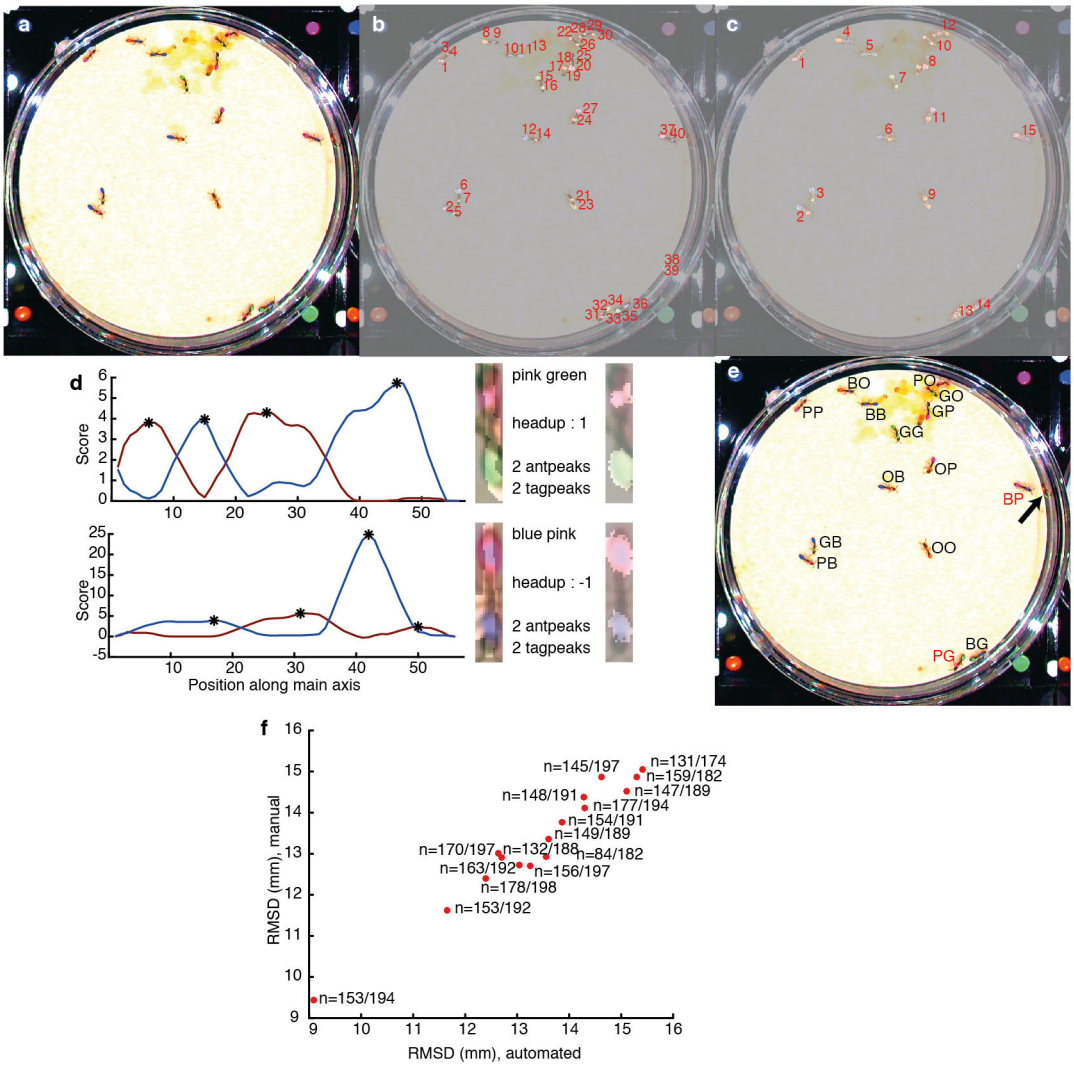
$$P_{ij,t} = \frac{s_{j,t}^{\eta}}{s_{j,t}^{\eta} + \theta_{ij}^{\eta}}$$

For large values of η we recover a deterministic behavior, 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 behaviors. Fourth, and finally, 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 to start performing.

To analyze 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].

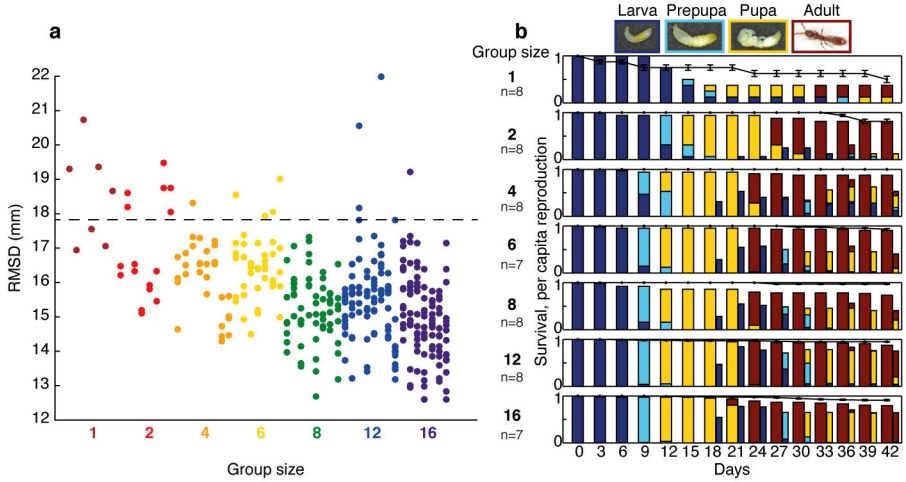
Extended Data



Extended Data Figure 1. Ant detection algorithm.

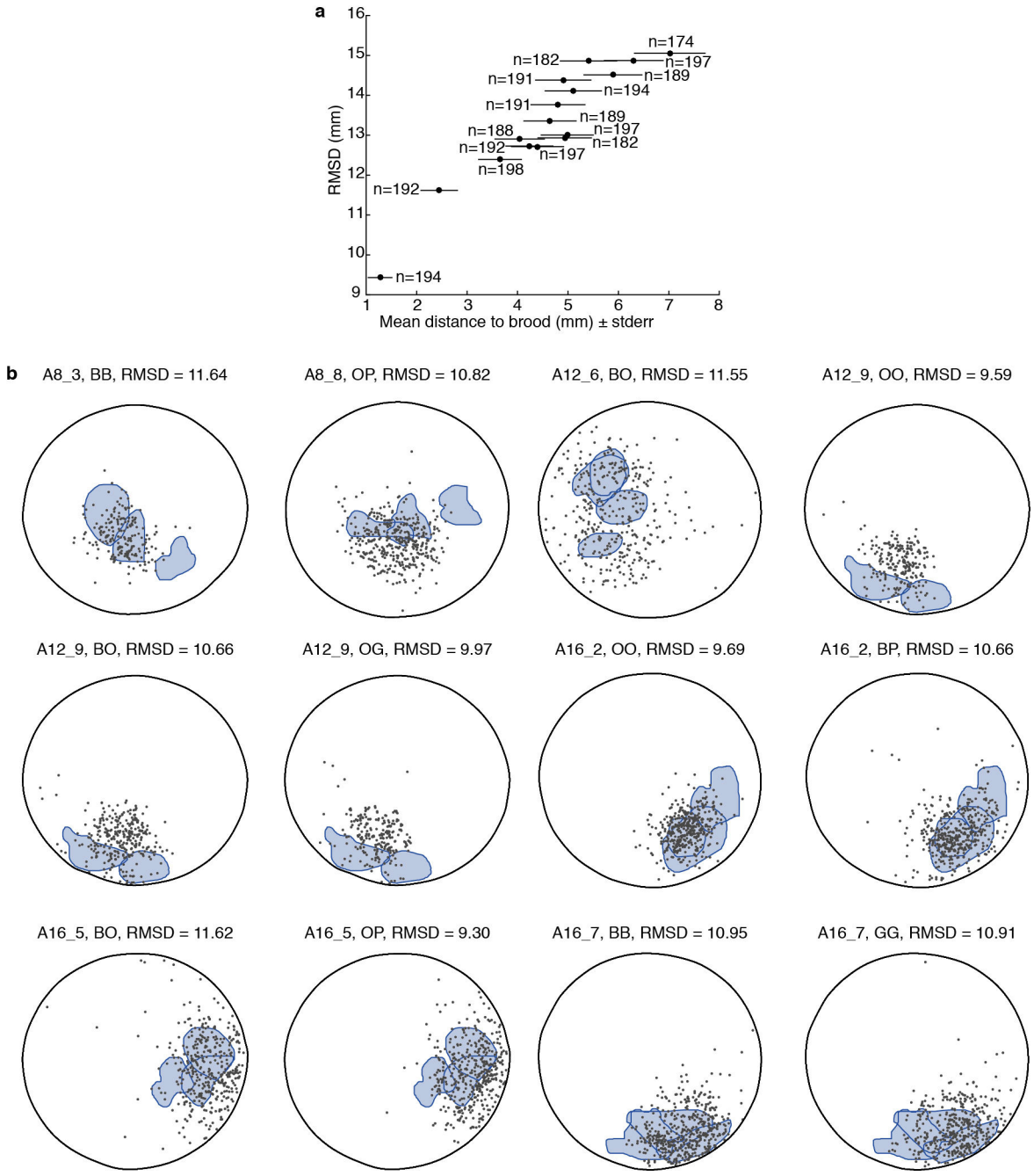
a, Example cropped frame showing one 16-worker colony following image correction. **b,** Color tag detection. The highlighted, numbered zones are image regions containing pixels that were assigned a high probability for tag colors (green, blue, orange, or pink) by a

Bayesian classifier. **c,** Candidate ant detection. The highlighted, numbered zones correspond to contiguous regions containing color tags and pixels that were assigned a high probability for ant color (i.e. cuticle) by the classifier. **d,** Candidate ant orientation. Candidate ants are aligned using the segment connecting the two color tags, and oriented (head down vs. head up) based on 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 color IDs (thorax-abdomen, G: green, B: blue, O: orange, P: pink). Ants shown as examples in d are labeled in red. All shown assignments are correct, but one ant is missed (arrow). This panel is identical to Fig. 1a. **f,** Correlation between RMSD calculated from automated vs. manual assignments for one 16-worker colony. RMSD was computed from a subset of frames in the brood care phase. n=(# automated detections)/(# manual detections). Pearson's r=0.95, p<0.001.



Extended Data Figure 2. RMSD and fitness in genotype B.

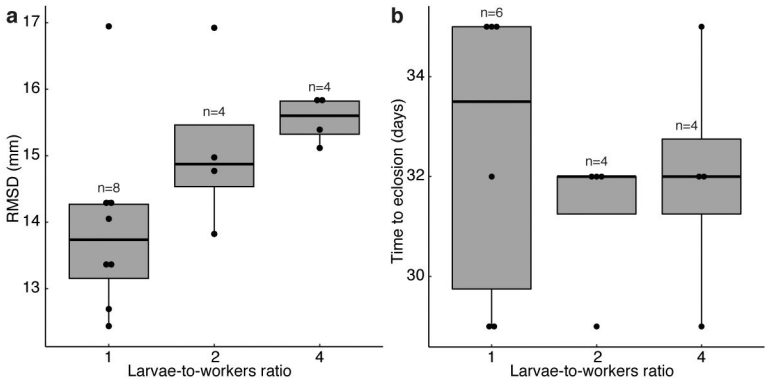
a, Individual RMSD values for all workers of genotype B. Ants from the same colony are vertically aligned. The dashed line represents the expected RMSD assuming a uniform distribution of an ant's positions. **b,** The dynamics of brood development as a function of group size in genotype B. Proportion of the brood in successive developmental stages (colors) in colonies of sizes 1–16. Wide and narrow bars indicate first and second brood generations, respectively. Black line: worker survival (mean ± s.e.).



Extended Data Figure 3. RMSD and spatial fidelity to the nest.

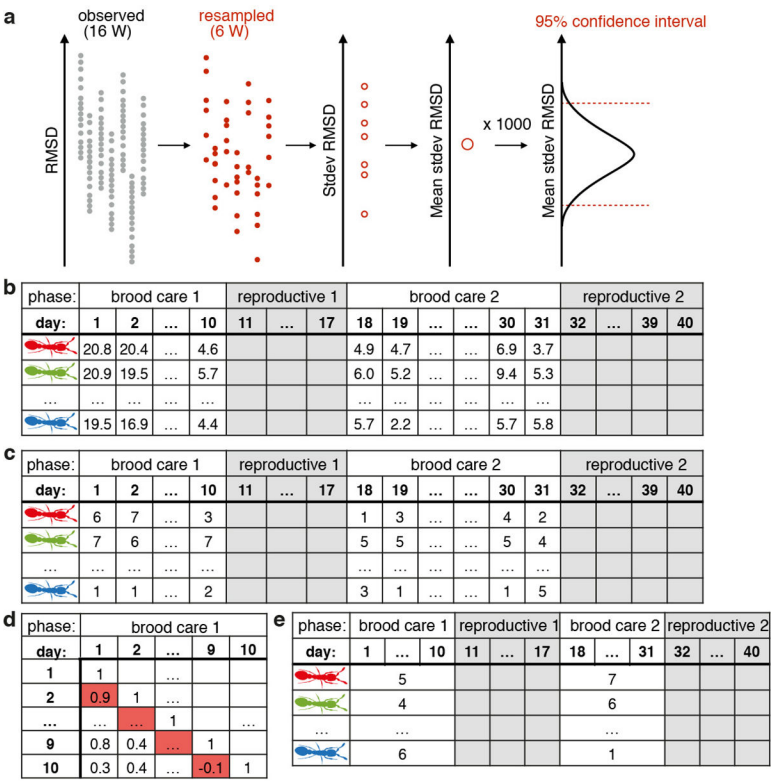
a, Correlation between individual RMSD and individual distance to the brood (mean ± s.e.) over the first brood care phase in one colony of 16 workers (Spearman's r=0.93, p=0; n=16 ants). Behavioral traits based on 209 manually tracked frames. Sample sizes (*n*) indicate the number of frames in which 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 RMSD (RMSD<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

(e.g. A8_3 is the 3rd replicate colony of genotype A and size 8), ant identity (e.g. BO for blue-orange) and individual RMSD. In each colony, the brood pile was manually annotated every 3 days (i.e. if the brood care phase lasted 9 days, 3 brood piles zones were manually annotated; zones could overlap or not, depending on how much the brood pile moved).



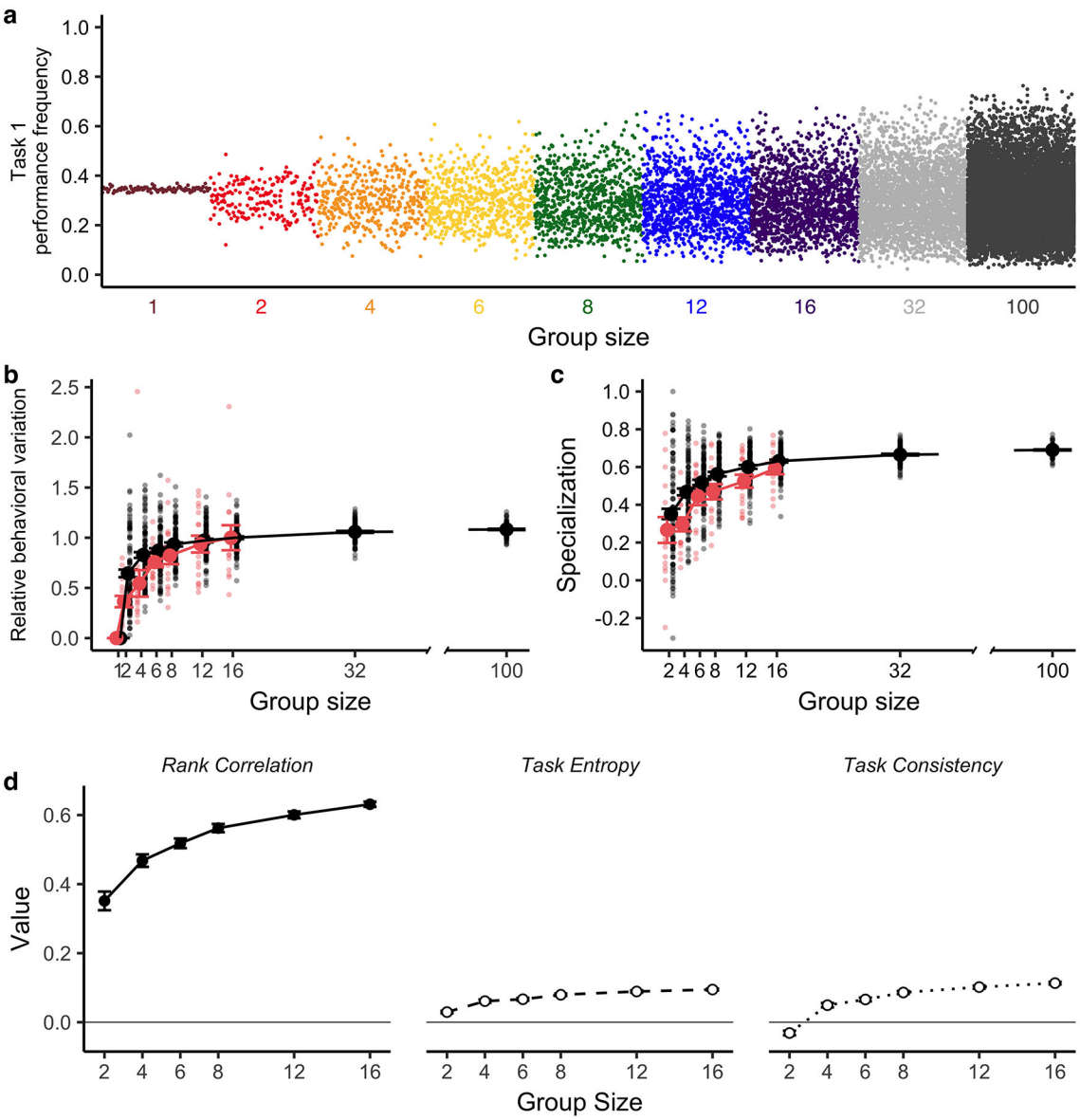
Extended Data Figure 4. Effect of the larvae-to-workers ratio on behavior and brood developmental time.

The number of workers was constant at 4, while 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 RMSD increased with the larvae-to-workers ratio (log-transformed RMSD: χ^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: χ^2=0.17, p=0.68). Sample sizes indicate the number of colonies in which at least one larva reached adulthood. In both panels, boxplots represent the median (thick horizontal line); the lower and upper hinges correspond to the first and third quartile, respectively. The upper (respectively, lower) whiskers extend from the upper (respectively, lower) hinge to the largest (respectively, smallest) value no further than 1.5 * IQR from the hinge, where IQR is the inter-quartile range.



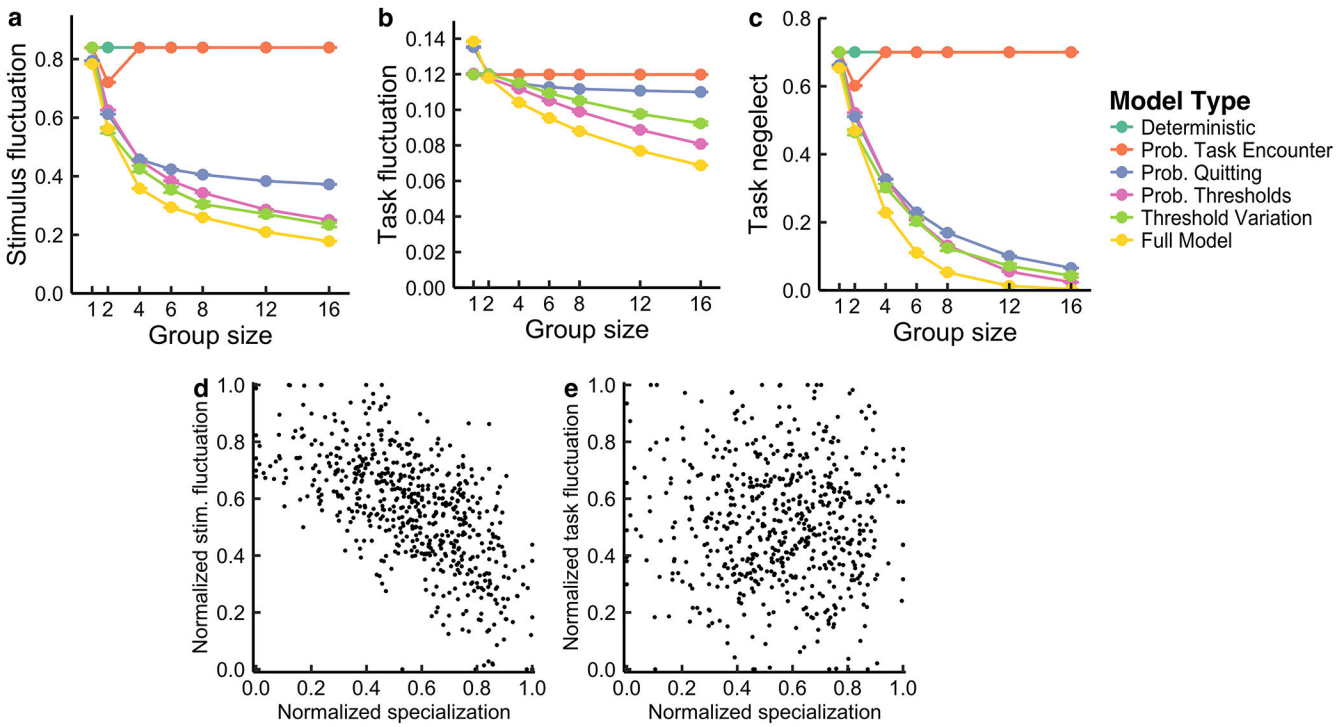
Extended Data Figure 5. Behavioral analyses methods.

a, Resampling scheme. 95% confidence intervals were generated by resampling individual RMSD values from one colony of size 16 at a time. Example shown here: generation of confidence intervals for behavioral variation (stdev RMSD) in colonies of size 6. The same method was used to generate confidence intervals for mean colony behavior (mean RMSD). **b-e,** Computing specialization. **b,** Daily individual RMSD values in one colony of size 8. **c,** Daily individual RMSD-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 behavioral consistency. **e,** Mean-RMSD ranks per brood care phase used to compute long-term behavioral consistency.



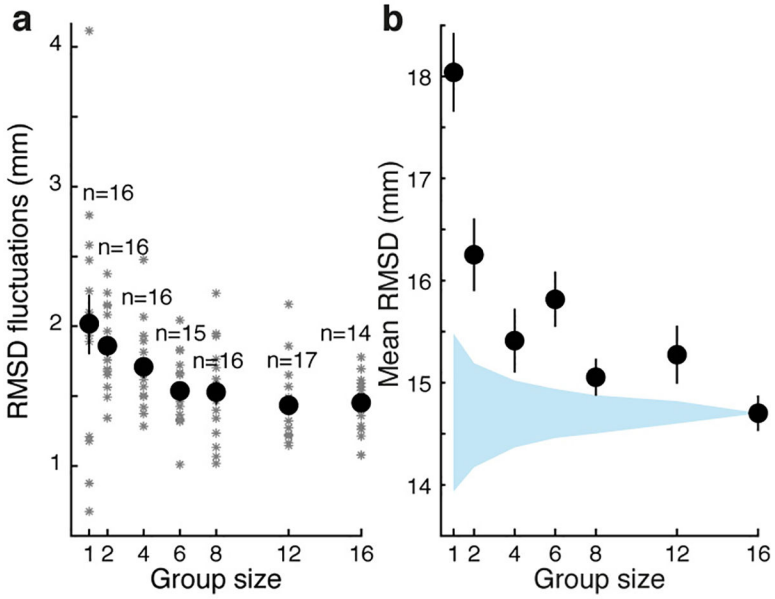
Extended Data Figure 6. Behavior of the fixed threshold model.

100 replicates were simulated per group size. Parameterization: $m=2$, $\eta=7$, $\mu=10$, $\sigma=0.1$, $\tau=0.2$, $\delta=0.6$, 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; ants from the same colony are vertically aligned. **b,** Behavioral 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, the line represents the mean value (± s.e.) across all 100 replicates for each group size. Model output is in black, experimental data is in red in (b)-(c). **d.** Mean values (± s.e.) of the rank correlation, task entropy, and task consistency metrics across all 100 replicates at each group size.



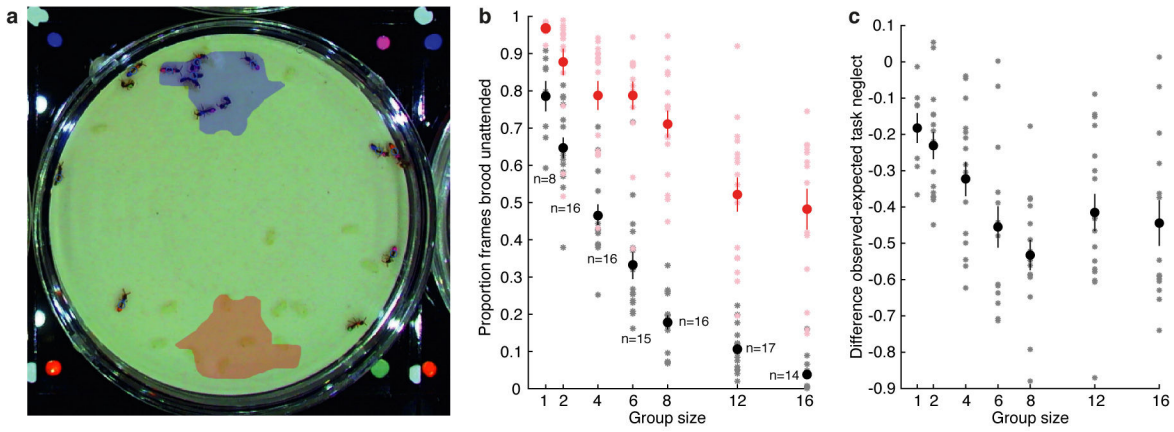
Extended Data Figure 7. The effect of stochasticity and specialization on proxies for fitness.

100 replicates were simulated per group size. Parameter settings for deterministic model are in Supplementary Methods; departures from deterministic model parameters are as in Extended Data Figure 6. **a,** Short-term (single time step) stimulus fluctuations averaged across both tasks; 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; shown across group sizes and for all models. **c,** Task neglect averaged across both tasks; shown across group sizes and for all models. In (a)-(c), points represent the described averages, further averaged (± s.e.) across n=100 replicate colonies of a given size. **d-e,** Relationship between specialization and (d) short-term stimulus fluctuations and (e) short-term fluctuations in task performance frequency, in the full model when controlling for group size. Each point represents one simulated colony.



Extended Data Figure 8. Behavioral homeostasis increases with group size.

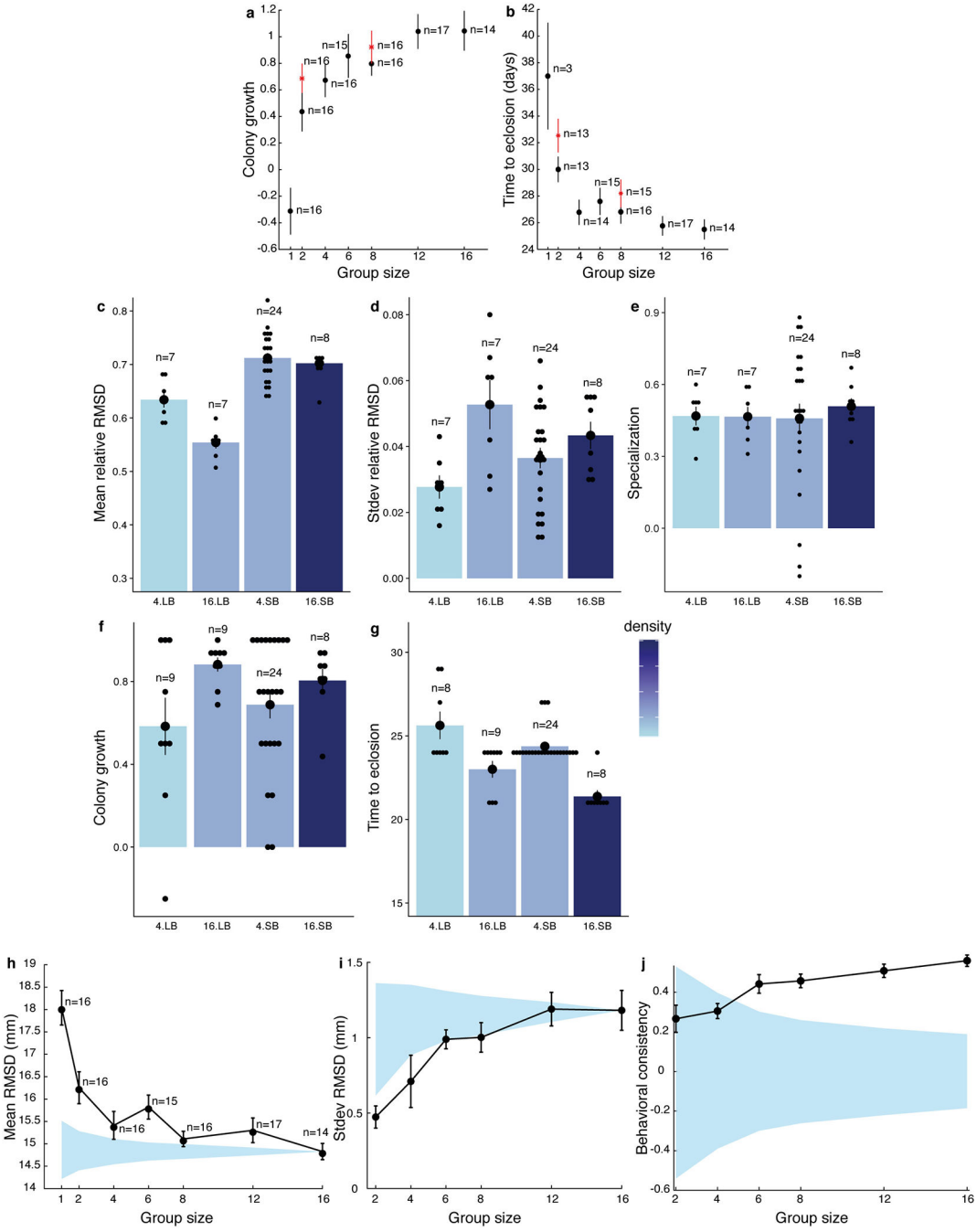
a, Day-to-day fluctuations in colony mean RMSD (mean ± s.e.) decrease with group size (LRT: χ^2=21.30, p=3.93*10^{-6}). Asterisks represent colony-level data. **b,**, Mean spatial fidelity increases with group size. Black: colony mean RMSD as a function of group size (mean ± s.e.). Blue: 95% confidence intervals under the null-hypothesis of no group-size effect on individual behavior, generated by resampling individuals from colonies of size 16 (Extended Data Fig. 5a). Sample sizes are as in (a). Data are pooled for genotypes A and B in both panels.



Extended Data Figure 9. Task neglect.

a, Manually annotated nest area (blue) and control area (red) generated by rotating the nest area by 180° around the center of the Petri dish. **b,** Task neglect (mean ± s.e.) 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.) decreases with group size (χ^2=13.36, p=2.57*10^{-4}). The difference

between observed and expected task neglect, as a function of group size. Sample sizes are as in (b). In (b)-(c), data are pooled for genotypes A and B and asterisks represent colonies.



Extended Data Figure 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.). Colony growth was unaffected by paint-marking (χ^2=2.71, p=0.10), the interaction of paint-marking with group size (χ^2=0.31, p=0.58), or

the interaction of paint-marking with genotype (χ^2=0.17, p=0.68). **b,** Larval time to eclosion in colonies of unmarked ants (mean ± s.e.). Time to eclosion of larvae was increased by paint-marking of the workers (square-root transformed time to eclosion: χ^2=8.98, p=0.003), but paint-marking did not interact with group size (χ^2=0.09, p=0.77) or genotype (χ^2=0.22, p=0.64). **c-g,** Effects of density on behavior 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.) was affected by group size (χ^2=6.49, p=0.01), box size (χ^2=38.46, p=5.6*10^{-10}), and density (group size:box size: χ^2=6.76, p=0.009). **d,** Behavioral variation (mean ± s.e.) was affected by group size (χ^2=7.44, p=0.006), but not by box size (χ^2=0.08, p=0.77) or density (group size:box size: χ^2=3.50, p=0.06). **e,** Behavioral consistency (mean ± s.e.) was not affected by group size (χ^2=0.03, p=0.87), box size (χ^2=0.22, p=0.64) or density (group size:box size: χ^2=0.02 p=0.88). Behavioral consistency was tranformed by (behavioral consistency+0.21)$^{1.5}$. **f,** Colony growth (mean ± s.e.) was affected by group size (χ^2=3.91, p=0.048), but not by box size (χ^2=0.04, p=0.85) or density (group size:box size: χ^2=1.00, p=0.32). Colony growth was transformed by (growth+0.4)$^{1.9}$. Thus, the impact 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.) was affected by group size (χ^2=35.74, p=2.26*10^{-9}) and box size (χ^2=10.45, p=0.001), but not by density (group size:box size: χ^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,** Colony mean spatial fidelity increases with group size. Black: mean RMSD (± s.e.) as a function of group size, after excluding individuals with 4 or more ovarioles. Blue: 95% confidence interval generated by resampling workers from 16-worker colonies (Extended Data Fig. 5a). **i,** Behavioral variation increases with group size. Black: standard deviation in RMSD per colony as a function of group size (mean ± s.e.), after excluding individuals with more than 3 ovarioles. 95% confidence intervals and sample sizes are as in a. **j,** Day-to-day rank consistency increases with group size. Black: mean RMSD rank correlation coefficients over consecutive days in the first brood care phase as a function of group size (mean ± s.e.), after excluding individuals with more than 3 ovarioles. Blue: 95% confidence intervals generated by randomizing daily ranks in each colony. Data pooled for genotypes A and B in panels (a), (b), (h), (i), and (j).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgements: We thank Asaf Gal for advice on data analysis, Ofer Feinerman and Molly Liu for contributions to the tracking algorithms, Stanislas Leibler, Zak Frentz, and David 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 #8.

References:

1. Queller DC Cooperators since life began. Q. Rev. Biol 72, 184–188, (1997).
2. Nowak MA, Tarnita CE & Wilson EO The evolution of eusociality. Nature 466, 1057–1062, (2010). [PubMed: 20740005]
3. Berdahl A, Torney CJ, Ioannou CC, Faria JJ & Couzin ID Emergent sensing of complex environments by mobile animal groups. Science 339, 574–576, (2013). [PubMed: 23372013]
4. Morand-Ferron J & Quinn JL Larger groups of passerines are more efficient problem solvers in the wild. Proc. Natl. Acad. Sci. U.S.A 108, 15898–15903, (2011). [PubMed: 21930936]
5. Waters JS, Holbrook CT, Fewell JH & Harrison JF Allometric scaling of metabolism, growth, and activity in whole colonies of the seed-harvester ant *Pogonomyrmex californicus*. Am. Nat 176, 501–510, (2010). [PubMed: 20735259]
6. Dornhaus A, Powell S & Bengston S Group size and its effects on collective organization. Annu. Rev. Entomol 57, 123–141, (2012). [PubMed: 21888521]
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. U.S.A 115, 756–761, (2018). [PubMed: 29311307]
8. Fewell JH & Harrison JF Scaling of work and energy use in social insect colonies. Behav. Ecol. Sociobiol 70, 1047–1061, (2016).
9. Jeanson R, Fewell JH, Gorelick R & Bertram SM 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 JL & Anderson C Emergent polyethism as a consequence of increased colony size in insect societies. J. Theor. Biol 215, 363–373, (2002). [PubMed: 12054843]
11. Oldroyd BP & Fewell JH Genetic diversity promotes homeostasis in insect colonies. Trends Ecol. Evol 22, 408–413, (2007). [PubMed: 17573148]
12. Jeanson R & Weidenmuller A Interindividual variability in social insects - proximate causes and ultimate consequences. Biol. Rev. Camb. Philos. Soc 89, 671–687, (2014). [PubMed: 24341677]
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 Sociaux 53, 20–26, (2006).
15. Oxley PR et al. The genome of the clonal raider ant *Cerapachys biroi*. Curr. Biol 24, 451–458, (2014). [PubMed: 24508170]
16. Sendova-Franks AB & Franks NR 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 DM Dynamics of Task Switching in Harvester Ants. Anim. Behav 38, 194–204, (1989).
18. Mersch DP, Crespi A & Keller L Tracking individuals shows spatial fidelity is a key regulator of ant social organization. Science 340, 1090–1093, (2013). [PubMed: 23599264]
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). [PubMed: 28569746]
20. Crall JD et al. Spatial fidelity of workers predicts collective response to disturbance in a social insect. Nat. Commun 9, 1201, (2018). [PubMed: 29615611]
21. Weidenmuller 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, Mendez V, Andrade JS, Jr. & Espadaler X Variability in individual activity bursts improves ant foraging success. J. Royal Soc. Interface 13, (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. Royal Soc. B 263, 1565–1569, (1996).

24. Pacala SW, Gordon DM & Godfray HCJ Effects of social group size on information transfer and task allocation. Evol. Ecol 10, 127–165, (1996).
25. Franks NR & Tofts C Foraging for work - how tasks allocate workers. Anim. Behav 48, 470–472, (1994).
26. Gorelick R, Bertram SM, Killeen PR & Fewell JH Normalized mutual entropy in biology: quantifying division of Labor. Am. Nat 164, 677–682, (2004). [PubMed: 15540157]
27. Teseo S, Chaline N, Jaisson P & Kronauer DJC Epistasis between adults and larvae underlies caste fate and fitness in a clonal ant. Nat. Commun 5, 3363, (2014). [PubMed: 24561920]
28. Crall JD 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). [PubMed: 23661762]
30. Holbrook CT, Kukuk PF & Fewell JH Increased group size promotes task specialization in a normally solitary halictine bee. Behaviour 150, 1449–1466, (2013).

Additional References (Methods)

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: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2008).
33. Dodds PS & Watts DJ Universal behavior in a generalized model of contagion. Phys Rev Lett 92, 218701, (2004). [PubMed: 15245323]
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).

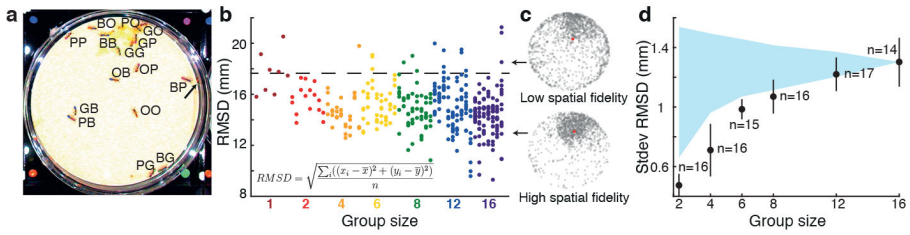


Figure 1. Behavioral variation as a function of group size.

a, Example frame showing the result of automated ant detection: 15 correct color-tag assignments (B: blue, G: green, O: orange, P: pink), 1 missed ant (arrow). **b,** Individual RMSD values for workers of genotype A. Ants from the same colony are vertically aligned. Insert: definition of RMSD, where (x_i, y_i) are the coordinates of the focal ant in frame i, $(\bar{x}, \bar{y})$ are the coordinates of the center of mass of the ant's spatial distribution throughout the brood care phase, and n is the number of frames in which the ant was detected. Dashed line: expected RMSD 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: center of mass. Arrows point to the corresponding ants in (b). Note that even workers with low spatial fidelity spend most time in the nest area. **d,** Behavioral variation increases with group size. Data pooled for genotypes A and B. Black: standard deviation in RMSD per colony as a function of group size (mean ± s.e.). Blue: 95% confidence intervals under the null-hypothesis of no group-size effect on individual behavior, generated by resampling individuals from colonies of size 16 (Extended Data Fig. 5a).

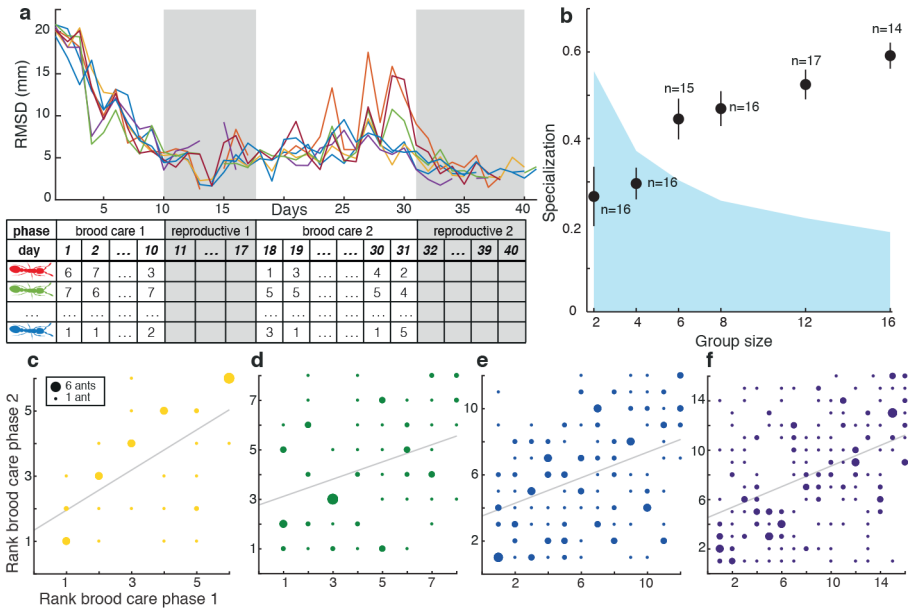


Figure 2. Specialization as a function of group size.

a, Daily individual RMSD in one colony (size 8, genotype B). The matrix shows a subset of the corresponding daily RMSD ranks. **b,** Specialization (mean ± s.e.) increases with group size. Black: mean day-to-day RMSD-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 behavioral rank across days, while 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 colonies of size: **c,** 6 ($r(42)=0.62$, $p=1.38*10^{-5}$, 95% CI: −0.32–0.32), **d,** 8 ($r(75)=0.35$, $p=0.002$, 95% CI: −0.23–0.24), **e,** 12 ($r(160)=0.40$, $p=1.31*10^{-7}$, 95% CI: −0.16–0.15), and **f,** 16 ($r(209)=0.44$, $p=3.42*10^{-11}$, 95% CI: −0.13–0.13). Circle diameter is proportional to the number of ants. Not shown: colonies of size 4 ($r(32)=0.08$, $p=0.68$, 95% CI: −0.42–0.42). Data pooled for genotypes A and B in (b)-(f).

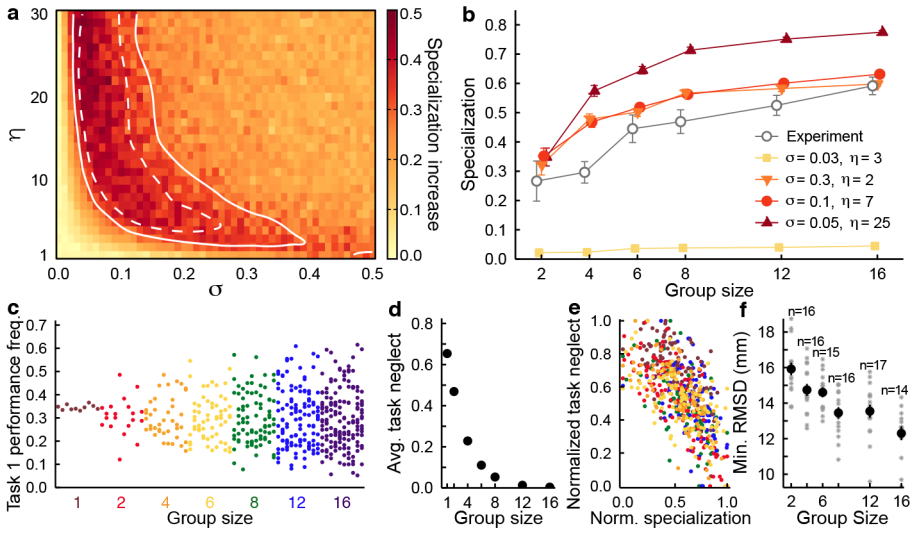


Figure 3. Results and predictions of the theoretical model.

100 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 = more deterministic threshold), and threshold variation, σ. Region between solid (lower bound) and dashed (upper bound) white contours encompasses simulated slope values approximated to be within 10% of experimental slope. **b,** Specialization (mean ± s.e.) as a function of group size for different parameter combinations of the theoretical model (color 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 (i.e. proportion of time, during a simulation run, in which a task went unperformed) across tasks. Points represent the average value (± s.e.) across all simulated colonies. **e,** Relationship between specialization and task neglect when controlling for group size. Each point represents one simulated colony; colonies are colored by group size as in (c). **f,** Minimum RMSD (mean ± s.e.), an empirical proxy for task neglect, decreases with group size (log-likelihood ratio test (LRT): χ^2=57.79, p=2.92*10^{-14}). Asterisks represent colony data. Data are pooled for genotypes A and B.

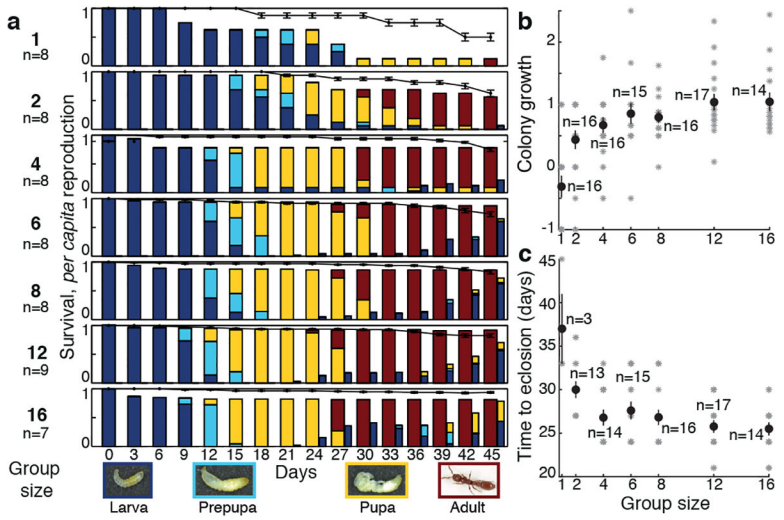


Figure 4. Fitness increases with group size.

a, The dynamics of brood development as a function of group size in genotype A. Mean proportion of the brood in successive developmental stages (colors); 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.). **b,** Colony growth (mean ± s.e.)—calculated as $(W_{end}\text{-}W_{start})/W_{start}$, where W_{end} and W_{start} are the number of live workers at the end and start of the experiment, respectively—increases with group size (LRT: χ^2=34.11, p=5.22*10^{-9}). **c,** Time to eclosion (mean ± s.e.) decreases with group size (LRT: χ^2=47.92, p=4.44*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 are pooled for genotypes A and B.