Uncovering Latent Behaviors in Ant Colonies

Mohamed Kafsi* Raphaël Braunschweig† Danielle Mersch† Matthias Grossglauser*
Laurent Keller† Patrick Thiran*

Abstract

Many biological systems exhibit collective behaviors that strengthen their adaptability to their environment, compared to more solitary species. Describing these behaviors is challenging yet necessary in order to understand these biological systems. We propose a probabilistic model that enables us to uncover the collective behaviors observed in a colony of ants. This model is based on the assumption that the behavior of an individual ant is a time-dependent mixture of latent behaviors that are specific to the whole colony. We apply this model to a large-scale dataset obtained by observing the mobility of nearly 1000 *Camponotus fellah* ants from six different colonies. Our results indicate that a colony typically exhibits three classes of behaviors, each characterized by a specific spatial distribution and a level of activity. Moreover, these spatial distributions, which are uncovered automatically by our model, match well with the ground truth as manually annotated by domain experts. We further explore the evolution of the behavior of individual ants and show that it is well captured by a second order Markov chain that encodes the fact that the future behavior of an ant depends not only on its current behavior but also on its preceding one.

1 Introduction

Many biological systems exhibit collective behaviors that endow them with immense adaptive advantages compared to more solitary species [4]. Together with humans, ant societies are probably the most widespread social organisms. They have colonized nearly all major landmasses, and their cumulated biomass tantamounts that of all vertebrates [13]. Moreover many behaviors and strategies previously thought unique to humans, like slave-making [2] and subsequent rebellion [1], as well as agriculture [11], use of antibiotics [13] and medical care [5] have also been developed by ants. However, in contrast to humans, ants succeed in all these deeds without need for hierarchical structure and top-down organization. Instead ant societies rely on local stimuli coupled to simple rules [4], but the exact nature of these stimuli and rules is still a mystery.

So far, behavioral analysis relied largely on small-scale hand-annotated datasets that describe the behavior of individual ants. Based on such data, researchers classify the behaviors and tasks associated with individual ants. Such manual annotations of individual behavior repertoires are highly subjective because they require that a human interprets a definition of a behavior and then quantifies it. Because analysis methods and interpretations vary extensively between researchers, studies differ in their results even if they analysed the same data [9, 14, 15].

The recent progress of tracking systems [10] now provides opportunities to collect and analyse large dataset that describe the behaviour of individual ants. With such a rich data comes the promise of a better understanding of the collective behaviors of an ant colony. However, this requires a shift in the approach taken when analysing behavioral data: we need to move from the traditional behavioral analysis to an objective approach that enables us to avoid the pitfalls of subjectivity.

In this paper, we take a data-driven approach to describe the behaviors observed in ant colonies. We propose a probabilistic model that enables us to uncover the collective behaviors in ant colonies, and to express the behavior of each ant as random combination of these collective behaviors. We apply this model to a large-scale dataset obtained by observing the mobility of nearly 1000 *Camponotus fellah* ants. Our results indicate that a colony typically exhibits three collective behaviors, each characterized by a specific spatial distribution and a level of activity. Moreover, the spatial distributions associated with the behaviors uncovered automatically by our model match well with the ground truth as manually annotated by domain experts. We further explore the evolution of the behavior of individual ants and show that it is well captured by a second order Markov chain that encodes the fact that the future behavior of an ant depends not only on its current behavior but also on its preceding one.

The remainder of the paper is structured as follows. In Section 2, we present the dataset that describes the

*EPFL, Lausanne, Switzerland.
†UNIL, Lausanne, Switzerland.

mobility of individual ants in different colonies. We then present our probabilistic model in Section 3, and explain how it captures the collective behaviors exhibited by a group of individuals. In Section 4, we apply this model to describe the behaviors exhibited in ant colonies and their evolution over time. We finally conclude the paper in Section 5.

2 Dataset

The dataset we use in this paper was collected by Mersch et al. [10]. They recorded, using an automated video tracking system, the activity of six colonies that are composed of a total of 956 *Camponotus fellah* ants. The set-up is divided in a nest area (Figure 1a) and a foraging area (Figure 1b) made of Plexiglas. Each area is rectangular and has on its short side an exit hole of 10 mm diameter that is connected to a tunnel that enables ants to move from one area to the other. Both areas are filmed from above with a high resolution camera (4560×3048 pixels) equipped with an enlarging lens and infrared light flash, to which ants are not sensitive.

In Figure 1, we show annotated images of the nest and foraging areas of colonies 4 and 18. The important areas of the nest, as delimited manually by Mersch et al. [10], are the brood pile (brown area), the rubbish pile (violet area), and the nest entrance (green area). The brood piles in the two colonies are very different as larvae are carried around by the ants of the colony. The foraging area, that we show in Figures 1b and 1d, contains a water source and food (liquid honey).

The queens and all workers are marked with a unique barcode that enables their identification in each image taken by the camera: their locations are estimated with a mean precision of 2.37 pixels (i.e., 0.14 mm, $0.8\% - 2\%$ of a *Camponotus fellah* ant). The continuous recording (2 images per second) results in a rich dataset that describes, with great accuracy and resolution, the mobility of each ant. Table 1 presents different statistics of the dataset.

No. of colonies	6
No. of ants	956
No. of days	11
Observations frequency	2 per second
No. of time steps	1,900,800
No. of observations	1,817,164,800

Table 1: Statistics of the dataset that describes the activity of 6 *Camponotus fellah* colonies.

3 Model

To an observer, detecting the collective behaviors that govern the colony of ants is challenging yet necessary in order to understand and describe the colony behavior. Hence, our model should enable us to uncover latent collective behaviors and describe the behavior of each individual ant as a function of these collective behaviors. In this work, we describe the behavior of an ant as a function of the following variables.

Location represents a location within the nest or foraging area, and is indexed by an integer $x \in \mathcal{X}$. We further distinguish the locations that are within the nest $x \in \mathcal{X}_n$ from those that are within the foraging area $x \in \mathcal{X}_f$.

Activity is a binary variable $a \in \{0, 1\}$ that indicates whether an ant is moving or not. Similarly to Mersch et al. [10], we declare an ant as inactive if it is still for more that 120 time frames (1 minute).

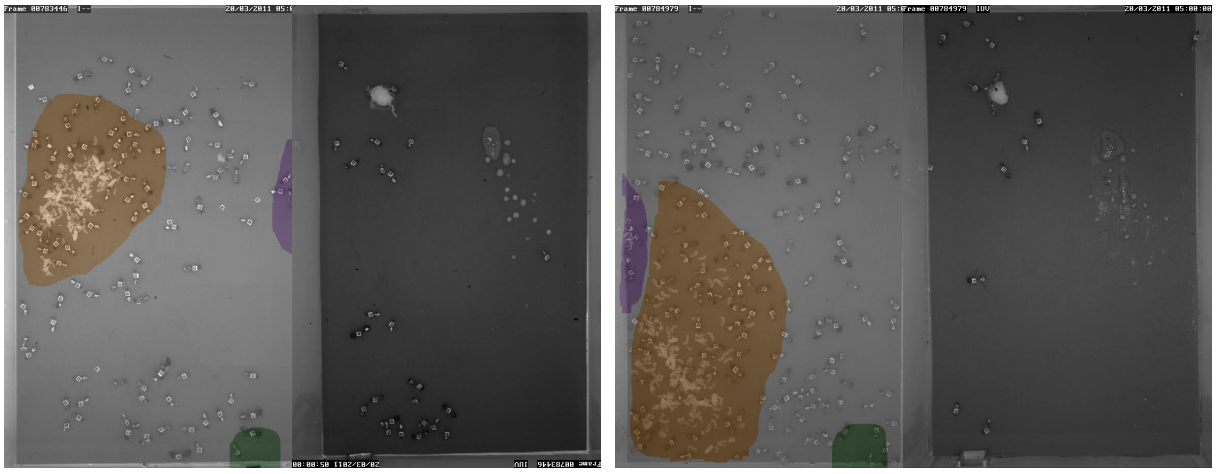
We describe the state of ant i at time t using two time-dependent stochastic processes $\{(X_i(t), A_i(t)), t \in \mathcal{T}\}$. The process $X_i(t) \in \mathcal{X}$ indicates the location of ant i at time t, and the process $A_i(t) \in \{0, 1\}$ indicates whether this ant is active at that time. As we are interested in describing macroscopically the geographical distribution of ants and their activity, we assume that

$$\begin{aligned} p(\boldsymbol{X}_i, \boldsymbol{A}_i) &= p(X_i(t_1), A_i(t_1), \dots X_i(t_n), A_i(t_n)) \\ &= \prod_{t \in \mathcal{T}} p(X_i(t), A_i(t)), \end{aligned}$$

where the joint distribution $p(X_i(t), A_i(t))$ describes probabilistically the state of the ant i at time t. As stated above, we assume that the individual behavior of an ant can be expressed as a random combination of K collective behaviors that are learnt from the behaviors of all individuals. This translates to the fact that we can express the joint distribution $p(X_i(t), A_i(t))$ with respect to a latent variable $z \in \{1, \dots K\}$ that indicates for each state (x, a) the behavior it was sampled from. More formally, we can write

$$\begin{aligned} p(X_i(t) = x, A_i(t) = a) &= \sum_{z=1}^{K} p(x, a|z) p(z|i, t), \\ &= \sum_{z=1}^{K} \theta_z(x, a)\, p(z|i, t). \end{aligned} \tag{3.1}$$

The K collective behaviors represent the collective dimensions of our model and are described by the shared multinomial distributions $\theta_k(x, a)$. The individual dimension is captured by the mixture coefficients $p(z|i, t)$



(a) Nest, colony 4. (b) Foraging, colony 4. (c) Nest, colony 18. (d) Foraging, colony 18.

Figure 1: We show the nest and foraging areas of colony 4 and colony 18. The important areas of the nest, as delimited by domain experts, are colored manually: The brood (brown), the rubbish pile (violet) and nest entrance (green).

that quantify to which extent the behavior of ant i at time t samples from the collective behavior k.

Given that we are interested in analyzing the individual ontogeny (mid-long term development) of ants while removing the variations that are due to the circadian rhythm (short term) of ants [7], we choose 24 hours as a temporal unit. Therefore, we assume that the behavior of an ant is stationary over a day i.e.,

$$p(z|i,t) = \pi_i(d(t), z).$$

Rewriting (3.1), we obtain

$$p(X_i(t) = x, A_i(t) = a) = \sum_{z=1}^{K} \theta_z(x,a)\, \pi_i(d(t), z),$$

where $d(t) \in \mathcal{D} = \{1, \ldots, 11\}$ is the day that corresponds to time t. Hence, the generative process, for ant i at time t, is as follows: (a) randomly adopt the behavior z according to the mixture coefficient $\pi_i(d(t), z)$, and then (b) randomly select a state (x, a) from the the joint distribution distribution $p(x, a|z)$. We also introduce a notion that will be useful for analyzing the evolution of the behavior of an ant: the dominating behavior $\beta_i(t)$ of ant i at time t is the behavior that maximizes the coefficient $\pi_i(t, z)$

$$\beta_i(t) = \operatorname*{argmax}_z \pi_i(t, z). \tag{3.2}$$

In other words, the behavior $\beta_i(t)$ is the most likely behavior given the data about ant i at time t.

In Table 2, we recapitulate the variables corresponding to an ant, as well as those relative to its state at time t.

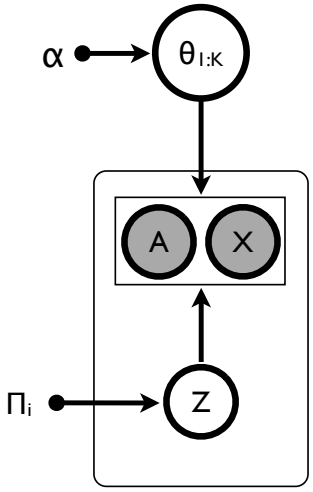


Figure 2: Graphical model representing the behavior of ant i at a given day. We assume that the distribution θ_k are drawn from a symmetric Dirichlet distribution parametrized by $\alpha = 0.1$.

3.1 Training The parameters of our model are the multinomial distributions θ_k associated with each canonical behavior and the mixture coefficients $\pi_i(d(t), z)$. In order to learn the model parameters that maximize the likelihood of data, we use the expectation-maximization (EM) algorithm [3]. This iterative algorithm increases the likelihood of the data by updating the model parameters in two phases: the E-phase and M-phase. The structure of our model enables us to de-

Definition	Domain	Explanation
t	$\mathcal{T} = \{t_1, \dots t_n\}$	Time step
c	$\mathcal{C} = \{4, 18, 21, 29, 58, 78\}$	Colony id
i	$\{\mathcal{I}_c, c \in \mathcal{C}\}$	Ant id
$\mathcal{X}$	$\mathcal{X} \subset \mathbb{N}$	Set of locations
$\mathcal{X}_f$	$\mathcal{X}_f \subset \mathcal{X}$	Set of locations in the foraging area
$\mathcal{X}_n$	$\mathcal{X}_n \subset \mathcal{X}$	Set of locations in the nest area
$d(t)$	$\mathcal{D} = \{1, \dots, 11\}$	Day associated with time t
k	$\{1, \dots K\}$	Canonical behavior id
K	$\mathbb{N}$	Number of canonical behaviors
$X_i(t)$	$\mathcal{X}$	Location of ant i at time t
$A_i(t)$	$\{0, 1\}$	Indicates whether ant i is active at time t
$\pi_i(t, z)$	$K-1$ simplex	Mixture coefficients of behavior z at time t for ant i
$\boldsymbol{\pi}_i(t)$	$-$	Vector of mixture coefficients $[\pi_i(t,1), \dots, \pi_i(t,K)]$ for ant i
$\beta_i(t)$	$\{1, \dots K\}$	Dominant behavior of ant i at time t
$\theta_k(x, a)$	$\mathcal{X} \times \{0, 1\}$	Multinomial distribution representing canonical behavior k

Table 2: List of the definition and domain of the variables relative to a an ant, as well as those describing its state at time t.

rive closed-form expressions for these updates. Due to lack of space, we omit to show our computations and refer the interested reader to Bishop [3] for the general EM algorithm and the approach to derive the updates for a particular mixture of distributions. In order to assign a non-zero probability to every state, we smooth the multinomial parameters: we assume that the distribution θ_k are drawn from a Dirichlet distribution parametrized by α. The Dirichlet distribution is a distribution of K-dimensional discrete distributions parameterized by a vector $\boldsymbol{\alpha}$ of positive reals. Its support is the closed standard simplex, and it has the advantage of being the conjugate prior of the multinomial distribution. In other words, if the prior of a multinomial distribution is the Dirichlet distribution, the inferred distribution is a random variable distributed also as a Dirichlet conditioned on the observed states. In order to avoid favoring one state over the others, we choose the symmetric Dirichlet distribution ($\alpha = 0.1$) as a prior.

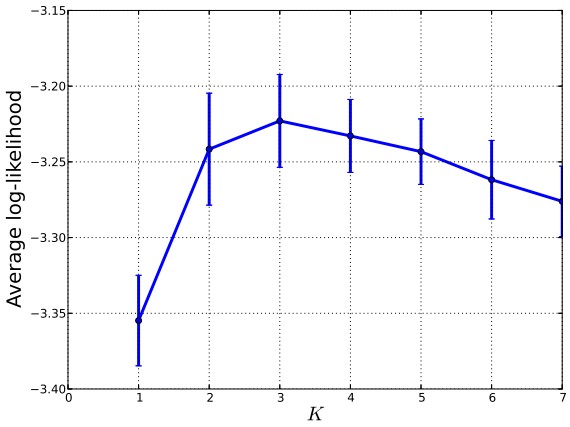


Figure 3: Average log-likelihood as a function of the number of collective behaviors K. These values are obtained by averaging the log-likelihoods of the 600 test sets obtained by randomly splitting the dataset of each colony.

4 Results

4.1 Collective behaviors

Finding the number of canonical behaviors In order to find the number of collective behaviors that explains the data the best, we repeat the following process 100 times for each colony c: We divide randomly the dataset of colony c in a training (95%) and a test set (5%), and train our model for different values of the number of collective behaviors $K \in \{1 \dots 7\}$.We then compute the log-likelihood of the test set, given the parameters learnt from the training set. We obtain the results shown in Figure 3 by averaging the log-likelihoods of the 600 (6 colony and 100 random splits) test sets obtained by randomly splitting the dataset of each colony. We observe that the number of behaviors that maximizes the average likelihood of the test sets is $K = 3$. We assume for the rest of the paper that we have three collective behaviors in each colony because this configuration best explains the observed data.

Describing the collective behaviors In order to be able to map the behaviors uncovered in different colonies between each other, we assign labels to these behaviors according to the probability of being in the nest.

- **Nest (N)** the behavior z that maximizes the probability of being in the nest $p(x \in \mathcal{X}_n|z)$.
- **Foraging (F)** the behavior z that maximizes the probability $p(x \in \mathcal{X}_f|z)$ of being in the foraging area.
- **Intermediate (I)** the remaining behavior.

Mobility In order to visualize the location distributions associated with the behaviors detected by our model, we plot in Figure 4, for colonies 4 and 18 and each behavior z, the top 1000 locations ranked according to the probability $p(x|z)$. These locations represent, for each behavior, the area where an ant that adopts this behavior would spend most of its time. Moreover, we show in Figure 6 the locations of each colony colored according to the mixture coefficients $p(z|x)$. We focus on Figure 4 and note that the top locations associated with the behavior N capture well the shape of the brood pile, as shown in Figure 1a. Moreover, the most likely locations associated with behavior F correspond to the entrance of the foraging area and food source (Figure 1b) where foragers spend most of their time.

In order to confirm these observations, we compare the mixture coefficients $\pi_i(d, z = N)$ for behavior N to the proportion of time an ant would spend in the brood pile as delimited by domain experts. In order to represent the structure of the ant colony, Mersch et al. [10] annotated manually the most important region in the nest, namely the brood pile (brown region in Figure 1a), and then measured the time each ant spends inside and outside this region. The purpose of these measurements is to associate ants with tasks in the colony, for example, nurses are the individuals that visit most often the brood pile. We denote by $\nu_i(d)$ the proportion of time ant i spends in the brood pile at day d, as measured by Mersch et al. [10]. This quantity provides us with a biological ground truth against which we can compare our results. In Figure 5, we plot this quantity as a function of the mixture coefficients $\pi_i(d, z = N)$ associated with behaviors N: $\nu_i(d)$, the proportion of time an ant spends in the brood pile increases with the mixture coefficient $\pi_i(d, z = N)$ associated with behavior N. More importantly, the fact that $\nu_i(d)$ is very close to $\pi_i(d, z = N)$ confirms that the spatial distribution associated with behavior N matches accurately the area of the brood pile, as delimited manually by domain experts. In order to confirm this observation, we measure the mean squared error between the mixture coefficients $\pi_i(d, z = N)$ and the actual proportion of time spent in the brood pile $\nu_i(d)$

$$\frac{1}{|\mathcal{D}| \sum_{c \in \mathcal{C}} |\mathcal{I}_c|} \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{I}_c} \sum_{d \in \mathcal{D}} (\pi_i(d, z = N) - \nu_i(d))^2.$$

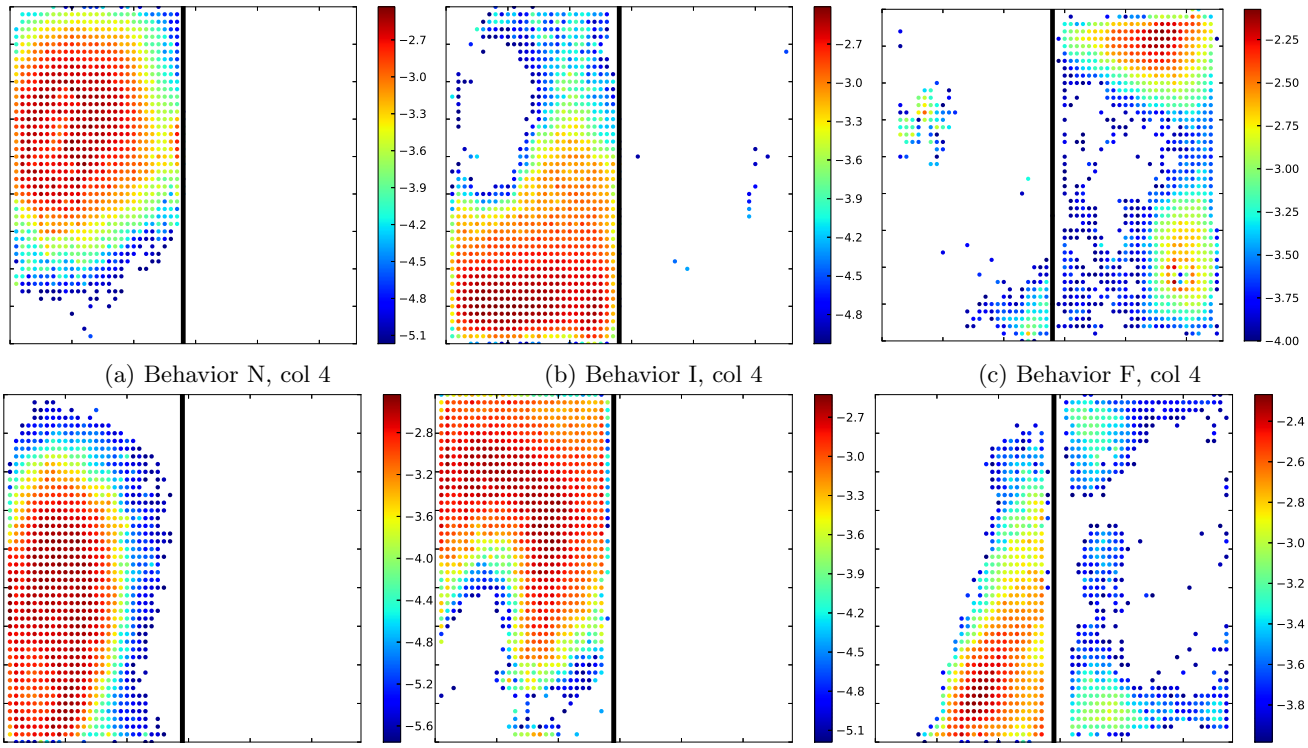
The mean squared error is equal to 0.03 which clearly confirms that the spatial distribution associated with the behavior N uncovered by our model accurately matches the spatial distribution of the nurses as defined by domain experts. Our approach would therefore enable biologists to automatically detect the *spatial fidelity* zones [8, 16, 10] in a colony without going through the tedious process of manual annotation.

	N	I	F
Colony 4	0.26	0.36	0.51
Colony 18	0.27	0.28	0.4
Colony 21	0.28	0.33	0.42
Colony 29	0.29	0.4	0.4
Colony 58	0.22	0.39	0.52
Colony 78	0.26	0.28	0.46
Average	**0.26**	**0.34**	**0.45**

Table 3: The probability of being active $p(a = 1|z)$ for each colony and each behavior.

Activity In Table 3, we show the probability of being active $p(a = 1|z)$ for each behavior and each colony. We notice that (a) inactivity prevails in the colony, and (b) the probability of being active increases as we move away from the brood pile. In fact, independently of the behavior adopted, an ant spends most of its time inactive: the probability of being active is $p(a = 1|z) = 0.35$ on average. However, this probability increases as the ants move away from the brood pile. For all colonies, the ants that adopt behavior N are less likely to be active than ants that adopt the intermediate behavior. The foragers are clearly the most active individuals, as their probability of being active is, on average, $p(a = 1|z) = 0.45$.

These results confirm the observations of Franks et al. [6] who monitored *visually* the behaviors of *Leptothorax acervorum* ants for nearly 12 hours in order to measure the proportion of time allocated to each activity. They found that nest workers spend 71.5% of their time resting, whereas foragers spend only 14.8% of their time resting. Even if findings are relative to a species of ants that is different from the one we are studying, they still support the validity of our observations about the relationship between the



(a) Behavior N, col 4 (b) Behavior I, col 4 (c) Behavior F, col 4

(d) Behavior N, col 18 (e) Behavior I, col 18 (f) Behavior F, col 18

Figure 4: We plot, for colony 4 and 18, the top 1000 locations of each behavior z ranked according to the probability $p(x|z)$. If we compare the spatial density associated with colony 4 to its configuration shown in Figures. 1a and 1b, we observe clearly that Behavior N is concentrated around the brood pile, whereas behavior F is concentrated around the water source and food.

behavior adopted and the activity level.

4.2 Individual Behaviors In our model, we represent the behavior of ant i at day d by the vector of mixture coefficients $\boldsymbol{\pi}_i(d)$, which is a point in the 2-simplex. We show in Figure 7 a scatter plot of the mixture coefficients $\boldsymbol{\pi}_i(d)$ of the ants associated with colony 4: Most of these points are concentrated on the edges of the triangle that represents the 2-simplex, which suggests that the majority of ants adopt a mixture of two behaviors. For example, a point that is close to the edge from N to I corresponds to an ant whose behavior is a combination of behaviors dominated by the collective behaviors N and I, but exhibits weak signs of behavior F. In such a representation, the evolution of an ant's behavior over time is represented as a *behavioral* trajectory (sequence of points) on the 2-simplex, similar to those shown in Figure 8. In order to gain a better understanding about the evolution of ant behavior, we study the evolution of the dominating behavior over the 11 days of observation. Hence, we assume that the sequence of dominating behaviors $\beta_i(1), \ldots, \beta_i(11)$ reflect the evolution of ant i.

Trajectory	**Empirical frequency**
F	17%
N	16%
I	11%
N → I → N	5%
I → N	4%
I → F → I	4%
I → F	3%
I → N → I → N	3%
F → I → F	3%
I → N → I → N → I	3%
I → N → I	2%
Others	29%

Table 4: Top trajectories ranked according to their empirical frequency.

Typical trajectories We show in Table 4 the most frequent behavioral trajectories from which we have removed self-transitions, i.e., we keep only transitions between different dominating behaviors. A large proportion of the ants has the same dominating behav-

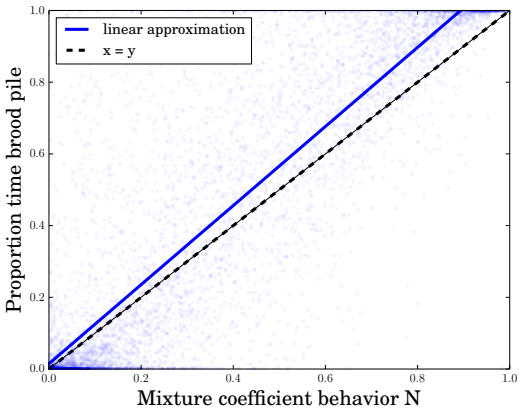


Figure 5: A scatter plot that represents the proportion of time in the brood pile $\nu_i(d)$ vs. the mixture coefficient $\pi_i(d, z = N)$ for behavior N. We show the linear approximation (intercept = 0.03, slope = 1.1) of the relation between these two variables and remark that it is very close to the line $x = y$, which represents a perfect match between the results of our model and the ground truth as measured by domain experts.

ior for the whole observation period. In fact, the probability $p(\beta_i(11) = \ldots = \beta_i(1))$ that an ant keeps the same dominating behavior throughout the observation period is equal to 0.43. However, this probability depends greatly on the initial dominating behavior, as the probability of keeping the initial behavior is 0.49, 0.26 and 0.62 for behaviors N, I and F, respectively. By observing the most frequent trajectories, we notice that ants that witness more than one change of dominating behavior then to revert to their preceding dominating behavior. For example, the most frequent trajectory that includes a change in the dominating behavior corresponds to an ant that starts with the dominating behavior N, evolves to behavior I before reverting back to behavior N.

Modeling behavior evolution We assume that the sequence of dominating behaviors experienced by an ant is a realization of a Markov chain [12] whose order is unknown. To find the order of the Markov chain that best explains the data, we repeat the following procedure 1000 times: We randomly divide the dataset in a training set (90% of the ants) and a test set (10% of the ants). Then, we learn, using a maximum-likelihood estimator on the data from the training set, Markov chains with different orders. Finally, we compute, for each Markov chain, the log-likelihood of the test set; the higher the likelihood, the higher the predictive power of the model. We obtain the average log-likelihood associated with each order of the Markov chain by computing the log-likelihood average over the 1000 random splits of the data. We show the results in Table 5: The second order Markov chain maximizes the average log-likelihood of unseen data. This suggests that, in order to predict the behavior of ant i at day $d + 1$, we have to take into account not only its current dominating behavior $\beta_i(d)$, but also its preceding dominating behavior $\beta_i(d - 1)$.

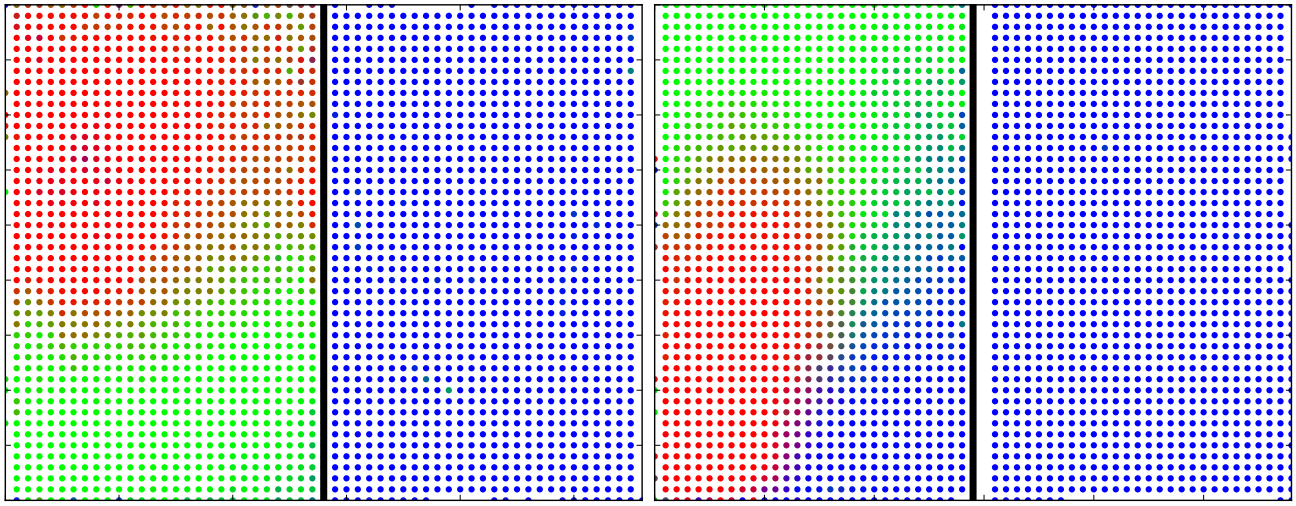
	Average log-likelihood
Random	-0.48
MC $\mathcal{O}(0)$	-0.47
MC $\mathcal{O}(1)$	-0.26
MC $\mathcal{O}(2)$	**-0.25**
MC $\mathcal{O}(3)$	-0.27
MC $\mathcal{O}(4)$	-0.30
MC $\mathcal{O}(5)$	-0.32

Table 5: The average log-likelihood of the test set is computed, for each method, over 1000 random splits of the dataset. We also indicate this average log-likelihood if we predict behaviors uniformly at random.

State (probability)	N	I	F
N $\rightarrow$ N (0.3)	0.9	0.09	0.01
I $\rightarrow$ I (0.27)	0.08	0.84	0.08
F $\rightarrow$ F (0.26)	0.01	0.07	0.92
N $\rightarrow$ I (0.04)	0.45	0.52	0.03
I $\rightarrow$ N (0.04)	0.66	0.31	0.03
I $\rightarrow$ F (0.04)	0.03	0.37	0.58
F $\rightarrow$ I (0.04)	0.03	0.56	0.41
F $\rightarrow$ N (0.007)	0.58	0.22	0.2
N $\rightarrow$ F (0.006)	0.51	0.14	0.35

Table 6: The second order Markov chain that models the evolution of the dominating behavior. We rank the states of the Markov chain according to their probability and show, for each state, the distribution over the next behaviors.

In Table 6, we show the second order MC learnt from our dataset. We rank the states according to their probability and show, for each state, the distribution over the next dominating behavior. The results confirm that preceding behaviors have an influence on future behaviors. For example, imagine that we observe an ant whose current behavior is I and we are interested in the probability that its next behavior is F. The preceding behavior of this ant influences this probability, which it



(a) Colony 4 (b) Colony 18

Figure 6: The color of each location x is a mixture of colors that is a function of the posterior probability $p(z|x)p(z)$. The colors red, green, and blue are mapped to the behaviors N, I and F, respectively.

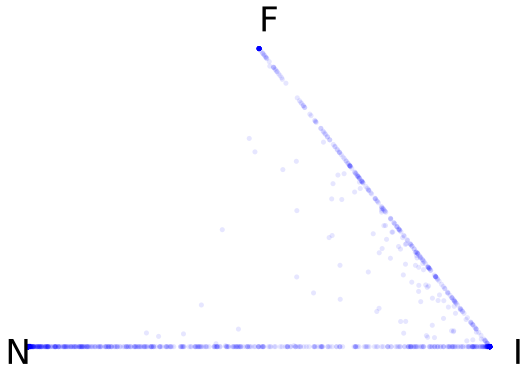


Figure 7: Each dot plotted in the simplex represents a mixture of behaviors (coefficient $\pi_i(d, z)$). These coefficients are concentrated on the edges of the simplex, which implies that an ant's behavior is dominated by two collective behaviors. A few ants expresses a balanced combination of the three collective behaviors.

is low if we know that this ant's preceding behavior is N or I (0.03 and 0.08 for N and I, respectively), whereas it increases significantly (0.41) if we know that this ant's preceding behavior is F.

5 Conclusion

In this work, we have modeled the behaviors of *Camponotus fellah* ants by analyzing large-scale digital traces that describe their mobility. We have proposed a probabilistic model that takes advantage of the correlation between the behavior of individual ants in order to uncover collective behaviors in a colony and to express the behavior of each ant as a time-dependent combination of these behaviors. The collective behaviors found by our model correspond to actual functional behaviors in ant colonies: The spatial distribution associated with them match well the spatial distribution as defined by domain experts. Moreover, our method quantifies to which extent each individual ant expresses a given collective behavior. We have further explored the evolution of the behavior of an individual ant and have shown that an ant exhibits an evolution pattern that is well captured by a second-order Markov chain: the future behavior of an ant depends not only on its current behavior but also on the preceding one.

Our work illustrates well how probabilistic models enable us to understand some of the mechanisms in biological systems: we take advantage of the correlation between individual behaviors in order to uncover latent collective behaviors and describe their properties.

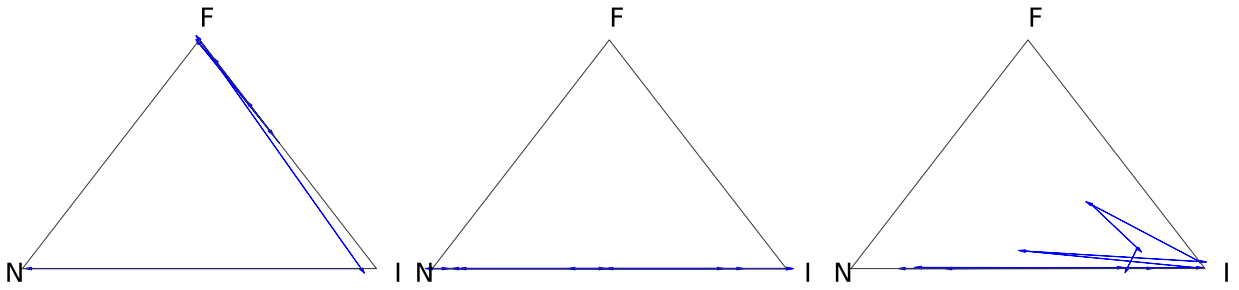


Figure 8: The evolution of the behavior of 3 ants, chosen from colony 4 and 18, are represented as a trajectory over the simplex.

References

[1] Alexandra Achenbach and Susanne Foitzik. First evidence for slave rebellion: Enslaved ant workers systematically kill the brood of their social parasite protomognathus americanus. *Evolution*, 2009.

[2] Hlldobler B and Wilson EO. *The ants*. Springer-Verlag, Berlin, 1990.

[3] C. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, New York, 2006.

[4] Scott Camazine, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau. *Self-Organization in Biological Systems*. Princeton University Press, 2001.

[5] Michel Chapuisat, Anne Oppliger, Pasqualina Magliano, and Philippe Christe. Wood ants use resin to protect themselves against pathogens. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1621):2013–2017, 2007.

[6] Nigel.R. Franks, Steve Bryant, Richard Griffiths, and Lia Hemerik. Synchronization of the behaviour within nests of the antleptothorax acervorum (fabricius)i. discovering the phenomenon and its relation to the level of starvation. *Bulletin of Mathematical Biology*, 52(5):597–612, 1990.

[7] Krista Ingram, Scott Krummey, and Michelle LeRoux. Expression patterns of a circadian clock gene are associated with age-related polyethism in harvester ants, pogonomyrmex occidentalis. *BMC Ecology*, 9(1):7, 2009.

[8] Jennifer M. Jandt and Anna Dornhaus. Spatial organization and division of labour in the bumblebee bombus impatiens. *Animal Behaviour*, 77(3):641 – 651, 2009.

[9] Steven A. Kolmes. Age polyethism in worker honey bees. *Ethology*, 71(3):252–255, 1986.

[10] Danielle P. Mersch, Alessandro Crespi, and Laurent Keller. Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science*, 340(6136):1090–1093, 2013.

[11] Ulrich G. Mueller and Christian Rabeling. A breakthrough innovation in animal evolution. *Proceedings of the National Academy of Sciences*, 105(14):5287–5288, 2008.

[12] J. R. Norris. *Markov Chains (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, 1998.

[13] Ted R. Schultz. In search of ant ancestors. *Proceedings of the National Academy of Sciences*, 97(26):14028–14029, 2000.

[14] Thomas D. Seeley. Adaptive significance of the age polyethism schedule in honeybee colonies. *Behavioral Ecology and Sociobiology*, 11(4):287–293, 1982.

[15] Thomas D. Seeley and Steven A. Kolmes. Age polyethism for hive duties in honey bees illusion or reality? *Ethology*, 87(3-4):284–297, 1991.

[16] A.B. Sendova-Franks and N.R.Franks. Spatial relationships within nests of the antleptothorax unifasciatus(latr.) and their implications for the division of labour. *Animal Behaviour*, 50(1):121 – 136, 1995.