REVIEW

The modelling cycle for collective animal behaviour

David J. T. Sumpter, Richard P. Mann and Andrea Perna*

Mathematics Department, Uppsala University, Uppsala, Sweden

Collective animal behaviour is the study of how interactions between individuals produce group level patterns, and why these interactions have evolved. This study has proved itself uniquely interdisciplinary, involving physicists, mathematicians, engineers as well as biologists. Almost all experimental work in this area is related directly or indirectly to mathematical models, with regular movement back and forth between models, experimental data and statistical fitting. In this paper, we describe how the modelling cycle works in the study of collective animal behaviour. We classify studies as addressing questions at different levels or linking different levels, i.e. as local, local to global, global to local or global. We also describe three distinct approaches—theory-driven, data-driven and model selection—to these questions. We show, with reference to our own research on species across different taxa, how we move between these different levels of description and how these various approaches can be applied to link levels together.

Keywords: collective animal behaviour; theoretical modelling; collective motion

1. INTRODUCTION

All biological systems have complex organization at different levels: molecules interact to form subcellular organules, cells organize together to produce tissues and so on. Even if we are aware that this complexity exists, most of the time we do not experience it directly with our senses. The interactions take place at spatial and temporal scales that are either too short or too long for us to grasp. But there is one level at which the complexity of biological interactions is immediately apparent to our eyes. It is the level of collective animal behaviour. When we observe a group of starlings coming together and flying in the evening light, a bank of fish attacked by a predator, a line of marching ants, we see the individual animals moving and interacting, and at the same time, we notice the appearance of another level of order of a different kind: the order of the flock, of the school, of the ant colony.

Understanding the behaviour of animal groups does not require a complete understanding of the behaviour of individual animals. Consider an everyday example of car traffic. Every rush hour, the patterns of congestion that form depend strongly on the density of vehicles, but weakly if at all on the origins or final destinations of the drivers. The same is true at all levels of physical, biological and social organization. Experiments and theoretical studies show that similar patterns of collective behaviour

emerge in systems as varied as humans, animals, tissue cells and microorganisms [1–3]. Often there is a characteristic dependence on the density of individuals, but the patterns created are independent of the specific identities of these individuals. We do not need to know the individual histories and motivations of hundreds of individuals to predict what the group as a whole will do.

On the other hand, individual motivations and interactions are often the object of interest in studies of animal behaviour. We are interested in understanding the basic behavioural cues followed by and the survival strategies adopted by individuals, even when our ultimate aim is to better understand group behaviour [4,5]. Investigation of these cues can take a wide range of forms, from looking at how optic flow can be used to measure distance [6] and avoid collisions [7] to testing hypotheses about optimal foraging decisions [8]. These and many other studies form the core of animal behaviour research, providing basic information about how individuals react to stimuli and interact with each other.

Creating a dichotomy between understanding the group through simplified assumption about individual behaviour and detailed quantification of how individuals interact can be an extremely useful approach in understanding collective animal behaviour. Ultimately, however, the research goal is to find mutual relations between these two levels of description. This point is especially relevant when looking at collective motion or spatial structures created by animal groups. In such situations, the behaviour of even large groups can be drastically altered by manipulating the strength of

*Author for correspondence (perna@math.uu.se).

interactions and the movement of just a fraction of individuals [9,10]. If we understand the nature of these interactions we can conceivably change them to control the group behaviour [11,12].

New technology, such as GPS [13], radio-frequency identification tags [14], three-dimensional image analysis [15,16] and radio tracking [17], means that we now have access to large quantities of data at both the individual and global levels. Precisely because collective animal behaviour occurs on temporal and spatial scales that we can relate to, means that it has become possible to automate tracking of interactions. This gives collective animal behaviour research a head start on many other areas of complex systems research, but it also means that we need to develop a clear framework within which to analyse, quantify and make use of these data.

In this study, we provide a basic classification of the different levels at which collective animal behaviour is studied. We also present three modelling methodologies that can be applied, which we call theory-driven, data-driven and model selection. We then use this framework to review case studies in phase transitions, collective structures and collective motion. Our focus is on the degree to which we can complete the modelling cycle for such systems. To what extent can we go from observing global patterns, to using simple models to describe plausible local interactions between individuals, to then empirically quantify these interactions, and finally predict and check consistency of the empirical rules with the global patterns?

2. METHODOLOGICAL APPROACHES

In some cases, individual behaviour is less important than understanding groups, and in other cases, groups can be largely ignored when studying individuals. It is worth then pausing to evaluate and classify the level of description on which different research projects operate. We use the following classification.

Local. Here, we study and make predictions about the movements or behaviour of individuals on the basis of the local configuration of the environment and other individuals around the animal. In analogy to physics, this can be thought of as the 'classical mechanics' approach.

Local to global. Here, we aim to directly link the behaviour of individuals to the classes of patterns produced by groups. This is the question of studying self-organization in biological systems [3]: to what extent can simple interactions produce interesting collective patterns? In almost all cases, the step from local to global is carried out with the help of a mathematical model, where simplifying assumptions have been made about how individuals interact. In physics, this approach is often referred to as 'kinetic theory' and sometimes as a subset of 'statistical mechanics'.

Global. The focus is entirely on the large-scale macroscopic pattern. Here, a theory about the evolution of a large-scale structure tries to predict how it will change in time or react to perturbations. This is a purely 'statistical mechanics' approach, where we describe ensembles of individuals.

Global to local. Here, we identify regularities in the group configuration or dynamics and use these to propose possible local rules of interaction between individuals. The global patterns place restrictions on the set of possible local interactions. This is often referred to as the 'inverse problem'.

In what follows, we give a number examples of studies falling in to each of these classes.

A defining feature of the study of collective animal behaviour is the interplay between experiment and theoretical models. Most research aims at either deriving or testing a model for how individuals interact to produce collective patterns. We classify modelling approaches in three ways.

Theory-driven approaches. Here, the process starts from models that in their first formulation have been only informally constrained by data. Theory-driven approaches are usually based on a model that predicts the large-scale properties from a set of local interactions (local to global). Only in a next stage are data used to support or refute the theory.

Data-driven approach. Under this approach, we start from the position of analysing and making sense of experimental or observational data. We try to identify consistent patterns, but only with informal use of models. This process is often either purely local, where experimental data are used to quantify the response of individuals to the stimuli that they encounter around themselves, or purely global, where the shape and structure of groups are quantified.

Model selection approach. This approach does not insist on the production of models either from data or from theory, but on their evaluation. Starting from a set of data and different models, it aims to quantify how closely each model approximates the data. The aim is to determine to what degree one model is preferable to another and search for better alternatives.

The theory-driven, data-driven and model selection approaches are complementary. They are all combined within a *modelling cycle*, where we move up and down between the local and global levels. This movement between levels imposes a structure on the modelling cycle. We now describe the character of this process, using several case studies.

3. PHASE TRANSITIONS

One of the original theory-driven models in the study of collective motion is the self-propelled particle model proposed by Vicsek and co-workers [18−20]. This model is local to global. Each particle adjusts its own direction to be nearer to the average direction of the particles within a local radius. At the global level, as the number of particles increases, there is a phase transition from disordered movement to collective order and organized motion, with spontaneous direction changes observed near the transition point. This prediction was confirmed in marching locusts placed in an experimental ring [21] and then later in field observations [22]. The prediction of the Vicsek model is 'universal', in the

sense that the same global pattern emerges from differing underlying rules [2] (although see Gregoire *et al.* [23] for a discussion of how rules determine outcome). Indeed, not just locusts, but a large number of animal species all reproduce the same transition from random to aligned motion [24–26].

The fact that individual locusts interact locally supports the basic assumptions of the Vicsek model [21,27]. However, the match between theory and experiments is primarily at the global level. These studies do not identify or quantify the cues and responses used by animals when interacting with each other or with their environment. In order to provide a local description of interactions, Bazazi *et al.* [28] manipulated the sensory inputs to locusts, finding that visual and tactile stimuli originating from behind the locust were those that resulted in collective motion. Marching bands of locusts and crickets may form because individuals are attempting to eat those in front of them and get away from those behind [29]. Protein hungry locusts and crickets cannibalize conspecifics, and collective motion may result from a 'forced march' [30,31]. On this basis of local mechanism (i.e. cannibalism) and a global outcome (i.e. collective motion), these researchers built a model to link these two levels. In the model, they hypothesized that hunger increases individual interaction strengths and showed that this assumption was sufficient to explain increased marching speed of hungry locusts [32].

While showing that a local-level model is sufficient to explain global-level data is essential, we are often interested in comparing multiple models and finding which one provides the best explanation. This point is particularly relevant when so many models can produce the same global-level outcome. It is here that model selection can be applied. In order to investigate how model selection might be used in collective animal behaviour, we looked at interactions of glass prawns [33]. Compared with the many kilometre wide marching bands created by locusts, the collective patterns made by these prawns are much less spectacular. The prawn interactions are also simpler, primarily mediated by tactile contacts. This simplicity makes them an ideal test bed for model selection.

The prawns were put in a ring-shaped arena (external diameter: 20 cm) and filmed for 6 min. At the global level, we found that the prawns exhibited a similar phase transition as in the Vicsek model (figure 1*a*). We also found that prawns increased their probability of changing direction when they had a prawn nearby (figure 1*b*). We then confirmed that the global pattern could be reproduced accurately by a variety of appropriately parametrized models; some, such as the Vicsek model, involved only local interactions, whereas others included memory of previous encounters and global interactions. To determine which model fits the data best, we first translated the rules of motion in each model into a probability that a focal prawn will change its direction of motion based on either the local or global configuration of other individuals around it. By assessing this probability against the actual observed events (either the focal prawn changes direction or it does not), we were allowed to calculate the *likelihood* for each model, i.e. the probability of

the complete data sequence of movements over the duration of the experiments conditioned on the model being examined. This process also gave us the estimate of the model parameters. To check the consistency of these parametrized models with patterns of direction changes at the global level, we then simulated the models with parameters estimated, and compared these with the collective motion of larger groups of prawns. Through this process, we could establish the existence of either non-local interactions or a memory of previous interactions. The likelihood of various models is ranked in figure 1*c*. The model with the highest likelihood is Markovian and has a very large ($\sim \pi$ radians) interaction zone. This interaction zone is a non-local effect—it extends far beyond the plausible sensory perception of the prawn itself and is inconsistent with the localized effect seen in figure 1*b*. Of those models whose likelihood exceeds the mean field model, only non-Markovian models allow for a *local* interaction range consistent with figure 1*b* ($\sim \pi/5$ radians). Simulation of these non-Markovian models confirms that they reproduce the group level alignment seen in figure 1*a*, which represents a necessary large-scale condition. The prawn study illustrates the importance of distinguishing sufficiency and necessity in model fitting. A large number of models were sufficient to explain the data, but the memory of interactions was necessary to retain consistency between descriptive levels and with biological intuition.

4. COLLECTIVE STRUCTURES

Sometimes, the behaviour of animal groups results in the formation of physical structures: the dams collectively built by beavers, the mounds of termites and the networks of galleries of naked mole rats [34]. One such structure is the pattern of trails collectively produced by animals and humans [3,35,36]. The formation of trails is in many respects analogous to a 'phase transition' from random to coherent state. In this case, the transition is not in alignment, but in the spatial distribution of the individuals. A simple illustration of such a transition is the double bridge experiment, where a colony of ants are presented with a branching bridge with the two branches leading to different food sources. As the flow of ants from the nest increases we go through two transitions. First, from a disorganized state where branches are chosen at random to a situation where a pheromone trail forms to just one of the food sources [37,38], then a transition from one to two trails as the food sources become crowded [39].

These transitions are captured by differential equation models of the number of ants going to each feeder as a function of flow rate. The underlying assumptions of these models are that the ants respond nonlinearly to pheromone concentration, and positive feedback determines which food source is selected [40–42]. Such models are global in the sense that they provide a mean-field model of trail formation: in the models pheromone concentrations are updated, depending on the average probabilities of choosing one branch or the other. A number of theory-driven local to global models
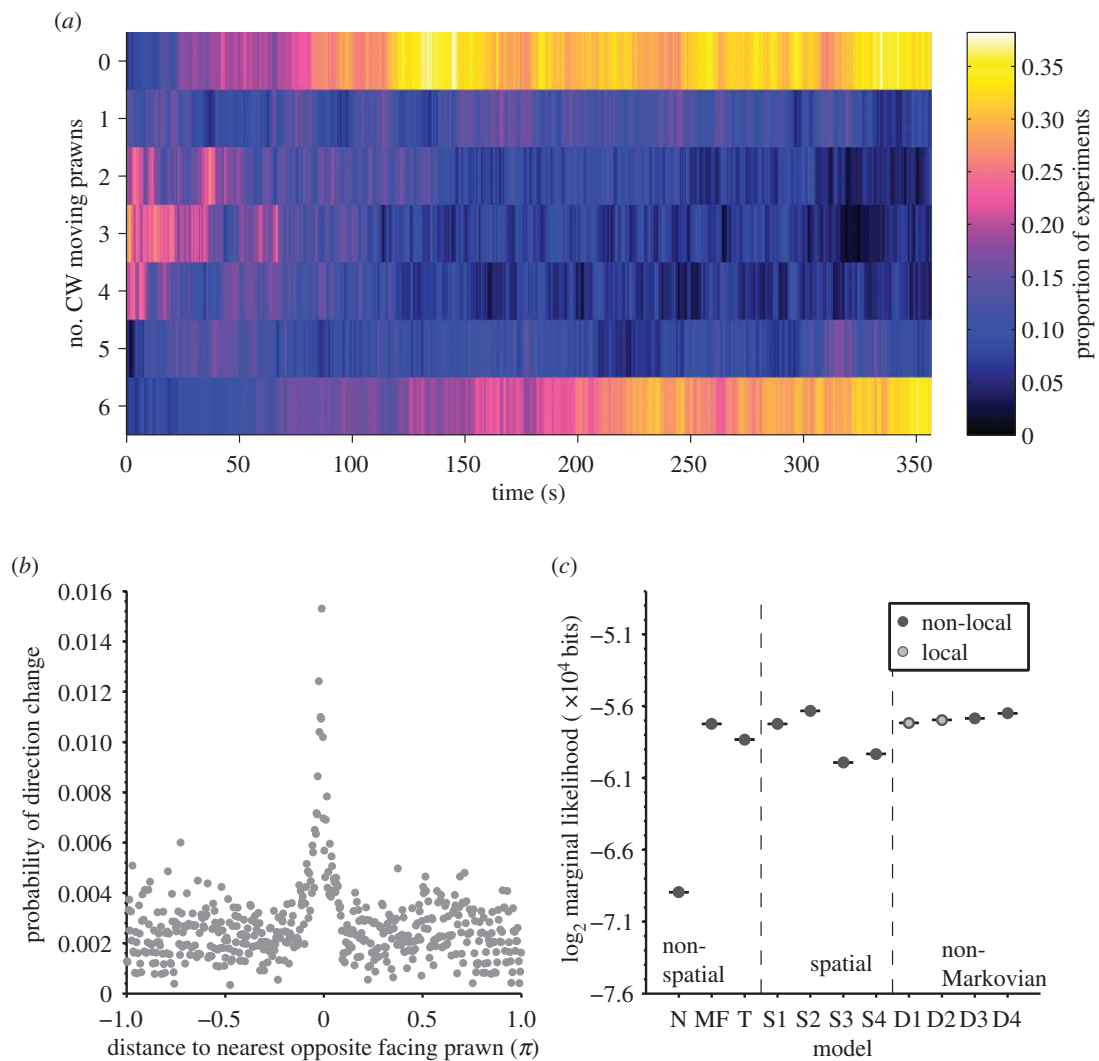
Figure 1. Analysis of collective behaviour in glass prawns. (*a*) The proportion of experiments with a given number of prawns travelling clockwise (CW) over time. This exhibits a phase transition from an initially disordered state, where most experiments have roughly half the prawns travelling CW, to aligned collective motion in either a CW or a counter-CW direction. (*b*) The probability that a prawn changes its direction of motion as a function of the position of the nearest conspecific travelling in the opposite direction. Prawns are more likely to change direction when close to another individual moving in the opposite direction. (*c*) Model selection favours either Markovian spatial models with large interaction radii (non-local interactions) or non-Markovian models with a local interaction. The model with the highest likelihood is a Markovian model with a large interaction zone (π radians). The only models with local interactions ($< \pi/2$ radians) are two non-Markovian models with memory effects, D1 and D2.

of trail formation have been proposed to explain the formation of trail networks [35,43–46]. These simulations reproduce the bifurcating structure seen in ant pheromone trails, again using unquantified, but biologically realistic, assumptions about how ants leave pheromone and react to it.

We adopted a data-driven approach to characterize the actual response of ants to pheromone concentrations [47]. We estimated pheromone concentrations by recording the passages of ants at all spatial positions in an experimental arena. These observations allowed us to infer a 'pheromone map' where we could look at correlations between pheromone concentrations and changes in speed or direction of movement of the ants (figure 2*a*). The ant turning angle α was a function of pheromone differences such that

$$\alpha = A \frac{L - R}{L + R + T_0}, \qquad (4.1)$$

where L and R are the concentrations of pheromone on the left and on the right side of the ant. A is a constant and T_0 is a threshold below which the ants cannot effectively detect pheromone. Such an individual response is the same as seen in human perception of sensory stimuli, such as sound loudness, image brightness, time, numerosity and weight [48]. This type of response, where the perception of a difference between two stimuli is inversely proportional to the magnitude of the stimuli, is known as Weber's Law [49].

Figure 2*b* shows how turning angle changes as a function of stimuli difference and total stimuli intensity, comparing experimental observations with equation (4.1). Here, the pattern of how response changes with pheromone concentration helps explain the outcome of the model of the transition from disorganized foraging to a choice of one of the feeders in the double bridge experiment. However, equation (4.1) is essentially linear, while a nonlinear response to pheromone was required
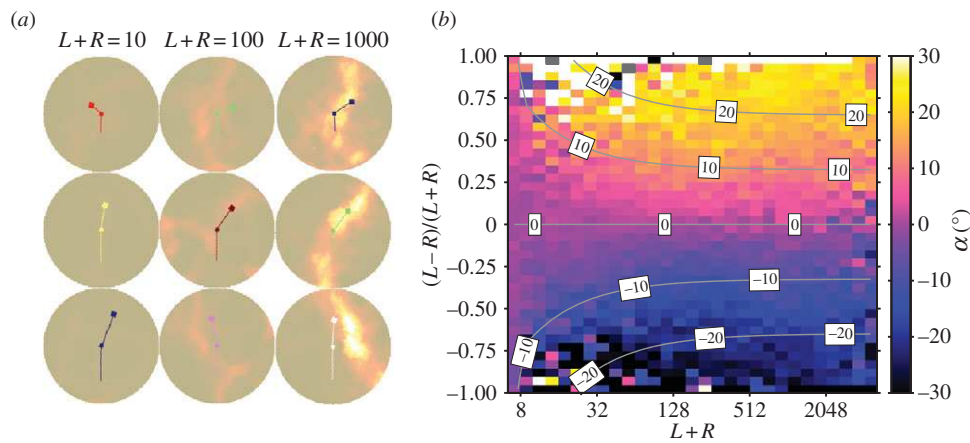
Figure 2. Ant response to pheromone concentrations. (*a*) Individual observations of ant movement on top of the 'pheromone map' produced by their own colony. Brighter colours indicate higher pheromone concentrations or, equivalently, more ant passages. The trajectory of the focal ant is summarized by its position at three consecutive time intervals of 400 ms each. Both the pheromone map and the ant's trajectory are rotated to be aligned together. The maps can be grouped together based on the amount of pheromone present in front-left (L) and front-right (R) sectors around the ant. (*b*) Turning angle (α) of the ant as a function of pheromone concentrations. The solid lines are the turning angles predicted by equation (4.1) with $A = 30.80$ and $T_0 = 10.53$. When the total pheromone concentration is low, ants' turning angle is less correlated with concentrations, but for higher concentrations the turning angle is proportional to $(L - R)/(L + R)$.

to produce this transition: trails can only be reinforced if the ants have a disproportionally higher probability to select the trail with higher pheromone concentration [42]. To reconcile these local and global levels, we showed that when integrating over repeated interactions between the ants and the pheromone trail, an appropriate nonlinearity appears, which explains the first of these transitions [47].

The data-driven approach allowed us to establish empirically how individual ants respond to pheromone and reconcile this with some aspects of their collective behaviour. However, when we implemented a spatially explicit simulation of trail formation in which ants follow Weber's Law, we found some differences between the exploratory pattern formed by the simulated ants and that formed by the real ants. In particular, simulated ants built more reinforcing loops that remain stable for indefinite time. These results suggest that some additional navigation ability such as path integration or response to other ants may be required to model exploratory network formation [50].

5. COLLECTIVE MOTION

The self-propelled particle models used to explain phase transitions can be extended to capture other aspects of collective motion. Models that include repulsion, attraction and alignment responses to neighbours reproduce many of the dynamic features of bird flocks and fish schools [51–58]. These models produce swarms, rotating tori, dynamic moving groups with continually changing internal structures as well as many other aspects of the movement of animal groups. Although these models may look reasonable at a global level, the question remains as to which (if any) of these models best captures what the animals actually do.

The first stage in linking these models to data is quantifying the global shape and dynamics of

flocks and schools. This is non-trivial, because the movements are highly complex and usually occur in three dimensions. Despite these challenges, basic information about the structure, shape and dynamics of starling flocks [15,59,60] and fish schools [61] has now been established. It may be possible to use some of these data to solve the inverse problem of going from global structure to local interactions. For example, the topological arrangement of neighbouring starlings is largely independent of flock density, suggesting that starlings interact with a limited set of nearest neighbours, rather than interacting with all neighbours within a fixed distance [15]. However, a local to global model in which individuals interact with neighbours in inverse proportional to distance between them also reproduces the same topological arrangement [55]. This modelling exercise again confirms that if we are really interested in evaluating and comparing alternative self-propelled particle models, we need at least some experimental information about how individuals interact locally.

Two recent studies have looked at local interactions between fish using a data-driven approach [62,63]. The rules of interaction were determined by identifying correlations between the relative positions and directions of individuals and the response of a focal individual, in terms of changes in direction or speed. By investigating how these responses vary as a function of the configuration of the neighbouring individuals, the interaction rule can be observed directly. For example, figure 3 plots the left–right component of the movement of the focal fish in response to the position of its neighbours (its 'lateral response' in figure 3*b*). The response is such that the focal fish generally turns towards the neighbours, and collision avoidance is mainly mediated through changes of speed [63].

The challenge in this approach lies in identifying which of many correlated variables is most important in predicting fish motion. For example, the acceleration of a focal fish is correlated both with the average
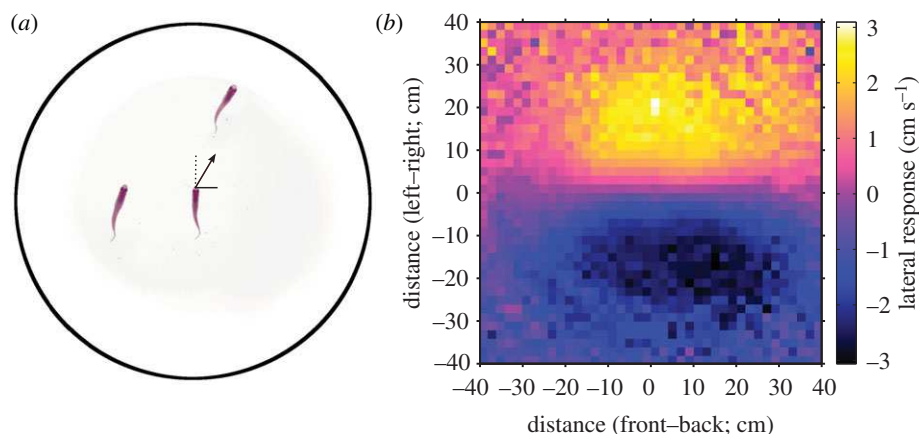
Figure 3. Fish response to their neighbours. (*a*) Snapshot of a shoal of mosquitofish seen from above. The focal fish (in the centre) responds to the neighbours by changing its speed and direction of movement. The lateral component of the response indicated by the solid line is plotted in (*b*) as a function of the position of the neighbours.

position of all its neighbouring fish and with the position of its nearest neighbour, which are themselves correlated variables. Which of these correlations is most relevant to predicting the fish's future movement? To answer this and similar questions, we fitted a sequence of models for each of the relevant factors. First, we determined how much of a fish's future speed and direction could be explained simply as a linear function of its recent direction and acceleration changes. Then, accounting for this autocorrelation, we fitted a neural network model of how the fish responded to its physical environment (i.e. tank walls). Then, accounting for both autocorrelation and physical environment, we fitted a further neural network model, this time for how the fish responded to its nearest neighbour. Up to this point, each additional model provided an improvement in prediction of fish movements. However, when we went on to fit a model of how the second and further away neighbours affected the fish's response, we found that these models provided no further improvements in terms of predicting the focal fish's movement. Only the position of the nearest neighbour was required to predict these movements.

While we use the word 'predict' in the previous paragraph to indicate whether or not including additional interactions improves our model fit, we have not applied a rigorous process of model selection to the data. Ideally, we would like to introduce one parameter at a time to the model and test, as we did using the Bayes factor with the prawns, whether this parameter improves predictive power. We have used a similar approach to infer the number and location of landmarks used by homing pigeons [64]. There we evaluated the information content of each element of the path, determined the most informative points and assessed successively whether each new proposed landmark contributed predictive value to the current model. For the fish, the large number of potential combinations of stimuli means that we cannot test them all. Ultimately, the sequence whereby we fitted models first for autocorrelation, then for walls and then for each of the neighbours is motivated by 'biological intuition' about how fish react to stimuli rather than by the formal rigour required by model selection. That said, in all

biological applications, we must rely on biological reasoning to select the possible subset of models before we apply mathematical tools. We will return to this point in §6.

The earlier-mentioned process gives a model that maps from configuration to response, either in the form of a large look-up table based on the observations or as a neural network with a large number of parameters. We could implement this mapping as a self-propelled particle simulation where at each time step, each particle evaluates its local neighbourhood and updates its speed and direction using this mapping. This model could then be used as a check as to whether global level statistics can be reproduced, helping us further evaluate whether the model matches the available data. Such a process has however some limitations compared with the self-propelled particle models discussed in the first paragraph of this section. These models have only a small number of parameters and can be stated in a handful of equations that relate explicitly to how animals respond in different situations. We can easily change the relative strengths of these responses and judge their effect on collective patterns. When our model is based on a large number of parameters inside a neural network, it is less clear on what theoretical basis we can manipulate interaction rules.

Rigorous fitting of self-propelled particle models to data would help scale up from individual level observations to predictions of global patterns. It is, in theory, possible to test the degree to which self-propelled particle models can capture the motion of real animals. Artificial flocking trajectories generated by a simulation can then, using least squares 'force matching', be fitted to the model which generated the data as well as alternative, incorrect models [65]. Only a small number of observations are required in order to reject the incorrect alternative models and accept the correct model. We have demonstrated a probabilistic Bayesian approach to parameter inference and model selection, allowing comparison of models with different numbers of parameters [66]. Models that over-fit the data were rejected by this method. Parameter inference in this framework provides not only the 'best-fit'
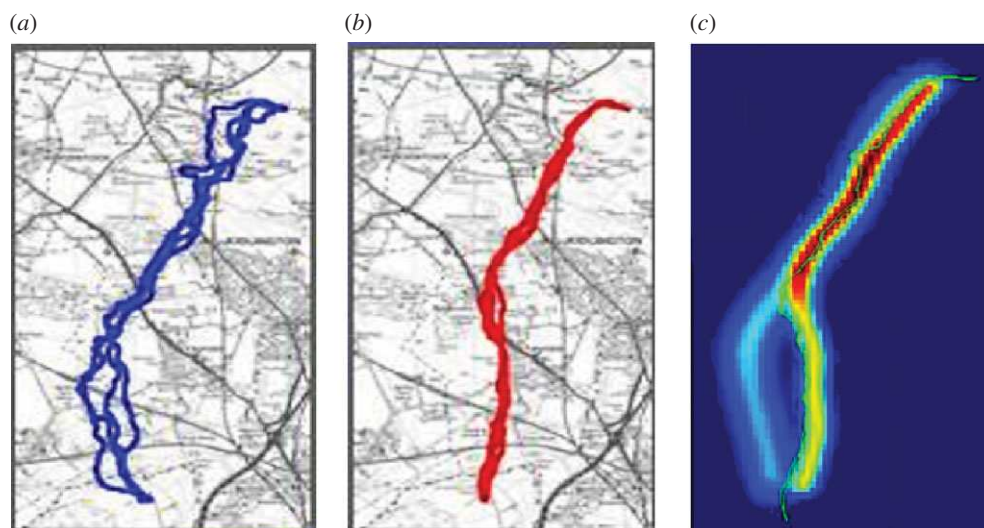
(a)   (b)   (c)

Figure 4. Prediction of paired pigeon flight paths from previous solo flights. (*a*) Pigeon 1 and (*b*) pigeon 2 are released repeatedly on their own from the same release site, developing idiosyncratic habitual routes back to the home loft. From the previous flights of each bird, a distribution of the likely spatial location of future flights can be determined [64]. (*c*) A weighted sum of these two distributions can predict the flight paths of each bird when they are released as a pair, with the two birds typically cooperating when the distributions overlap, and switching to either leader–follower behaviour or splitting when the distributions separate. Here, pigeon 2 leads pigeon 1 when the distributions diverge.

parameter values, but also how probable any alternative parameter sets might be.

An alternative and complementary research direction to fitting models to data is to focus purely on predicting future outcomes without an underlying model of interactions. For example, Mann [67] looked at how pairs of pigeons decided upon a joint flight route home when flying together, based on observations of their previous preferences when flying alone (figure 4*a,b*). A distribution of each individual's flight paths was inferred from multiple solo flights [64], which then gave the probability that each individual would fly along a particular route in subsequent flights. An assumption was then made that their joint behaviour would represent a blend of these individual preferences. Collective route selection was therefore modelled by a weighted mixture of the two birds' solo distributions, with no explicit 'rule of interaction'. Model selection was performed to determine what factors consistently affected the relative weights of the bird's own previous route and its partner's previous route. The data supported the hypothesis that a bird's degree of loyalty to its own route predicted how much it will tend to be influenced by its partner (figure 4*c*).

6. DISCUSSION

There is a particular methodology to approaching collective animal behaviour that emerges from the above examples. We start by observing a global pattern and formulate a theory-driven local to global model that is sufficient to explain certain aspects of this pattern. Then, the focus switches to experimental observations at the individual level. These reveal new details of the interactions between animals, which are incorporated into new, primarily data-driven, models. When we implement these data-driven models, we again test sufficiency against the global level data, hopefully identifying additional global level properties that our new model matches.

There are two potential outcomes of this process: either the data-driven model produces the global-level pattern or it does not. Both these outcomes imply further work. If the data-driven model fails to produce the global-level pattern, then we need to reassess the individual level data or perform further experiments in order to improve our model. Here, it may be useful to try and use the global-level pattern to infer local rules. In any case, the cycle between global and local should be repeated until more is learnt about why these levels cannot be reconciled. It is worth noting that failures to reproduce global patterns are much less widely reported in the literature. This is partly because scientists repeat this part of the cycle before publishing, and then only report the model that 'works'. Similarly, if no model works, then they might all be omitted from the paper. This, like other failures to report negative results, is an unfortunate but unavoidable limitation of written science.

While negative modelling results are under-reported, positive modelling results should not be over-interpreted. The fact that a model reproduces a collective pattern does not imply it is necessarily the best or even a good model of the data. Most scientists are aware of this distinction between sufficiency and necessity, but for collective animal behaviour, this point is particularly important. The importance arises as a direct consequence of the capability of theory-driven models to produce global patterns independent of model assumptions. In collective behaviour, most theory-driven models are 'sufficient' to produce realistic patterns of group-level behaviour of real animals, but none of the models is 'necessary': alternative models based on different assumptions often perform equally well and match the data to a similar precision. Thus,

the fact that one particular data-driven model also matches the data does not in itself constitute additional support for the model, over and above that already established by comparison with the data at an individual level. Only if the model is falsified by simulation (if the simulation does not reproduce the global pattern observed in the data) do we really gain additional understanding of the system as a whole.

It is here that model selection approaches can be useful. Once we have a set of potential models, we can use the Bayes factor to determine which gives the best explanation of the data, using the minimum number of necessary parameters. We applied such model selection to prawn collective motion and were able to establish the importance of non-local or non-Markovian interactions. A similar approach has been proposed by Weitz and co-workers, who suggest that models should be 'enunciated' and translated into quantitative terms in advance of comparison to individual and group level experiments [68]. In practice, model selection requires that we have already come quite far in understanding what is going on in a system. Computational limitations mean that we should have a limited number of models to test between. This necessarily implies that we must rely on a combination of data-mining—by which we mean plotting behavioural response data in various ways to identify patterns—and 'biological intuition' in order to produce a set of models to select between.

As well as requiring a set of models to select between, we need a criterion on which to make our selection. This may actually be easier to define at the local than at the global level. For example, we can measure the probability of turning left or right as a function of the animals' local environment and compare this with several models. The behaviour of individual animals can be further quantified in terms of their speed, their direction of movement, their probability to perform a specific action, etc. At the global level, the challenge is to quantify the form of a fish school or a flock of birds or a termite nest. Many properties can be measured, but it is difficult to discern between those that are actually relevant for understanding the group behaviour. As we collect more group level data, we envisage an increasing understanding of the relevant metrics in different cases.

Collective animal behaviour research is entering an exciting stage where elegant models are being confronted with detailed data of how animals move and interact. As a result, we have seen 'messier' models at an individual level, and new challenges as to how we explain the patterns these produce at the global level. Our hope is that through repeated revolutions of the modelling cycle we describe here, applied to a wide range of systems, we will provide a better understanding of how individual interactions produce collective dynamics.

REFERENCES

1 Sumpter, D. J. T. 2010 *Collective animal behavior.* Princeton, NJ: Princeton University Press.

2 Vicsek, T. & Zafeiris, A. In press. Collective motion. *Phys. Rep.* (doi:10.1016j.physrep.2012.03.004)

3 Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., Theraulaz, G. & Bonabeau, E. 2001 *Self-organization in biological systems.* Princeton, NJ: Princeton University Press.

4 Davies, N. B., Krebs, J. R. & West, S. 2012 *An introduction to behavioural ecology.* Oxford, UK: Wiley-Blackwell.

5 Krause, J. & Ruxton, G. 2002 *Living in groups.* Oxford, UK: Oxford University Press.

6 Srinivasan, M., Zhang, S., Altwein, M. & Tautz, J. 2000 Honeybee navigation: nature and calibration of the 'odometer'. *Science* **287**, 851–853. (doi:10.1126/science.287.5454.851)

7 Tammero, L. & Dickinson, M. 2002 The influence of visual landscape on the free flight behavior of the fruit fly *Drosophila melanogaster. J. Exp. Biol.* **205**, 327–343.

8 Galef, B. & Giraldeau, L. 2001 Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. *Anim. Behav.* **61**, 3–15. (doi:10.1006/anbe.2000.1557)

9 Couzin, I. D., Krause, J., Franks, N. R. & Levin, S. A. 2005 Effective leadership and decision-making in animal groups on the move. *Nature* **433**, 513–516. (doi:10.1038/nature03236)

10 Conradt, L., Krause, J., Couzin, I. D. & Roper, T. J. 2009 'Leading according to need' in self-organizing groups. *Am. Nat.* **173**, 304–312. (doi:10.1086/596532)

11 Halloy, J. *et al.* 2007 Social integration of robots into groups of cockroaches to control self-organized choices. *Science* **318**, 1155–1158. (doi:10.1126/science.1144259)

12 Faria, J., Dyer, J., Clément, R., Couzin, I., Holt, N., Ward, A., Waters, D. & Krause, J. 2010 A novel method for investigating the collective behaviour of fish: introducing 'robofish'. *Behav. Ecol. Sociobiol.* **64**, 1211–1218. (doi:10.1007/s00265-010-0988-y)

13 Biro, D., Sumpter, D. J. T., Meade, J. & Guilford, T. 2006 From compromise to leadership in pigeon homing. *Curr. Biol.* **16**, 2123–2128. (doi:10.1016/j.cub.2006.08.087)

14 Robinson, E. J. H., Richardson, T. O., Sendova-Franks, A. B., Feinerman, O. & Franks, N. R. 2009 Radio tagging reveals the roles of corpulence, experience and social information in ant decision making. *Behav. Ecol. Sociobiol.* **63**, 627–636. (doi:10.1007/s00265-008-0696-z)

15 Ballerini, M. 2008 Interaction ruling animal collective behaviour depends on topological rather than metric distance: evidence from a field study. *Proc. Natl Acad. Sci. USA* **105**, 1232–1237. (doi:10.1073/pnas.0711437105)

16 Straw, A. D., Branson, K., Neumann, T. R. & Dickinson, M. H. 2011 Multi-camera real-time three-dimensional tracking of multiple flying animals. *J. R. Soc. Interface* **8**, 395–409. (doi:10.1098/rsif.2010.0230)

17 Sword, G. A., Lorch, P. & Gwynne, D. 2005 Migratory bands give crickets protection. *Nature* **433**, 703. (doi:10.1038/433703a)

18 Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. 1995 Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**, 1226–1229. (doi:10.1103/PhysRevLett.75.1226)

19 Czirók, A. & Vicsek, T. 2000 Collective behavior of interacting self-propelled particles. *Physica A* **281**, 17–29. (doi:10.1016/S0378-4371(00)00013-3)

20 Czirok, A., Barabasi, A. & Vicsek, T. 1999 Collective motion of self-propelled particles: kinetic phase transition in one dimension. *Phys. Rev. Lett.* **82**, 209–212. (doi:10.1103/PhysRevLett.82.209)

21 Buhl, J., Sumpter, D. J. T., Couzin, I. D., Hale, J., Despland, E., Miller, E. & Simpson, S. 2006 From disorder

to order in marching locusts. *Science* **312**, 1402–1406. (doi:10.1126/science.1125142)

22 Buhl, J., Sword, G. A., Clissold, F. J. & Simpson, S. J. 2011 Group structure in locust migratory bands. *Behav. Ecol. Sociobiol.* **65**, 265–273. (doi:10.1007/s00265-010-1041-x)

23 Gregoire, G., Chate, H. & Tu, Y. 2003 Moving and staying together without a leader. *Physica D* **181**, 157–170. (doi:10.1016/S0167-2789(03)00102-7)

24 Becco, C., Vandewalle, N., Delcourt, J. & Poncin, P. 2006 Experimental evidences of a structural and dynamical transition in fish school. *Physica A* **367**, 487–493. (doi:10.1016/j.physa.2005.11.041)

25 Szabo, B., Szollosi, G. J., Gonci, B., Juranyi, Z., Selmeczi, D. & Vicsek, T. 2006 Phase transition in the collective migration of tissue cells: experiment and model. *Phys. Rev. E* **74**, 061908. (doi:10.1103/PhysRevE.74.061908)

26 Sokolov, A., Aranson, I. S., Kessler, J. O. & Goldstein, R. E. 2007 Concentration dependence of the collective dynamics of swimming bacteria. *Phys. Rev. Lett.* **98**, 158102. (doi:10.1103/PhysRevLett.98.158102)

27 Sumpter, D. J. T., Buhl, J., Biro, D. & Couzin, I. D. 2008 Information transfer in moving animal groups. *Theory Biosci.* **127**, 177–186. (doi:10.1007/s12064-008-0040-1)

28 Bazazi, S., Buhl, J., Hale, J. J., Anstey, M. L., Sword, G. A., Simpson, S. J. & Couzin, I. D. 2008 Collective motion and cannibalism in locust migratory bands. *Curr. Biol.* **18**, 735–739. (doi:10.1016/j.cub.2008.04.035)

29 Bazazi, S., Ioannou, C. C., Simpson, S. J., Sword, G. A., Torney, C. J., Lorch, P. D. & Couzin, I. D. 2010 The social context of cannibalism in migratory bands of the mormon cricket. *PLoS ONE* **5**, e15118. (doi:10.1371/journal.pone.0015118)

30 Simpson, S. J., Sword, G. A., Lorch, P. D. & Couzin, I. D. 2006 Cannibal crickets on a forced march for protein and salt. *Proc. Natl Acad. Sci. USA* **103**, 4152–4156. (doi:10.1073/pnas.0508915103)

31 Hansen, M. J., Buhl, J., Bazazi, S., Simpson, S. J. & Sword, G. A. 2011 Cannibalism in the lifeboat: collective movement in Australian plague locusts. *Behav. Ecol. Sociobiol.* **65**, 1715–1720. (doi:10.1007/s00265-011-1179-1)

32 Bazazi, S., Romanczuk, P., Thomas, S., Schimansky-Geier, L., Hale, J. J., Miller, G. A., Sword, G. A., Simpson, S. J. & Couzin, I. D. 2011 Nutritional state and collective motion: from individuals to mass migration. *Proc. R. Soc. B* **278**, 356–363. (doi:10.1098/rspb.2010.1447)

33 Mann, R. P., Perna, A., Strombom, D., Garnett, R., Herbert-Read, J. E., Sumpter, D. J. T. & Ward, A. J. W. 2012 Multi-scale inference of interaction rules in animal groups using Bayesian model selection. *PLoS Comput. Biol.* **8**, e1002308. (doi:10.1371/journal.pcbi.1002308)

34 Hansell, M. 2005 *Animal architecture.* New York, NY: Oxford University Press.

35 Helbing, D., Keltsch, J. & Molnár, P. 1997 Modelling the evolution of human trail systems. *Nature* **388**, 47–50. (doi:10.1038/40353)

36 Buhl, J., Hicks, K., Miller, E., Persey, S., Alinvi, O. & Sumpter, D. 2009 Shape and efficiency of wood ant foraging networks. *Behav. Ecol. Sociobiol.* **63**, 451–460. (doi:10.1007/s00265-008-0680-7)

37 Beekman, M., Sumpter, D. & Ratnieks, F. 2001 Phase transition between disordered and ordered foraging in Pharaoh's ants. *Proc. Natl Acad. Sci. USA* **98**, 9703–9706. (doi:10.1073/pnas.161285298)

38 Dussutour, A., Nicolis, S. C., Shephard, G., Beekman, M. & Sumpter, D. J. T. 2009 The role of multiple pheromones in food recruitment by ants. *J. Exp. Biol.* **212**, 2337–2348. (doi:10.1242/jeb.029827)

39 Dussutour, A., Fourcassie, V., Helbing, D. & Deneubourg, J. L. 2004 Optimal traffic organization in ants under crowded conditions. *Nature* **428**, 70–73. (doi:10.1038/nature02345)

40 Goss, S., Aron, S., Deneubourg, J. L. & Pasteels, J. M. 1989 Self-organized shortcuts in the Argentine ant. *Naturwissenschaften* **76**, 579–581. (doi:10.1007/BF00462870)

41 Deneubourg, J. L., Aron, S., Goss, S. & Pasteels, J. M. 1990 The self organizing exploratory pattern of the Argentine ant. *J. Insect Behav.* **3**, 159–168. (doi:10.1007/BF01417909)

42 Nicolis, S. C. & Deneubourg, J. L. 1999 Emerging patterns and food recruitment in ants: an analytical study. *J. Theor. Biol.* **198**, 575–592. (doi:10.1006/jtbi.1999.0934)

43 Watmough, J. & Edelstein-Keshet, L. 1995 Modelling the formation of trail networks by foraging ants. *J. Theor. Biol.* **176**, 357–371. (doi:10.1006/jtbi.1995.0205)

44 Edelstein-Keshet, L., Watmough, J. & Ermentrout, G. 1995 Trail following in ants: individual properties determine population behaviour. *Behav. Ecol. Sociobiol.* **36**, 119–133. (doi:10.1007/BF00170717)

45 Helbing, D., Schweitzer, F., Keltsch, J. & Molnár, P. 1997 Active walker model for the formation of human and animal trail systems. *Phys. Rev. E* **56**, 2527–2539. (doi:10.1103/PhysRevE.56.2527)

46 Couzin, I. D. & Franks, N. R. 2003 Self-organized lane formation and optimized traffic flow in army ants. *Proc. R. Soc. Lond. B* **270**, 139–146. (doi:10.1098/rspb.2002.2210)

47 Perna, A., Granovskiy, B., Garnier, S., Nicolis, S., Labédan, M., Theraulaz, G., Fourcassié, V. & Sumpter, D. J. T. 2012 Individual rules for trail pattern formation in Argentine ants (*Linepithema humile*). *PLoS Comput. Biol.* **8**, e1002592. (doi:10.1371/journal.pcbi.1002592)

48 Fechner, G. T. 1860 *Elemente der Psychophysik.* Leipzig, Germany: Breitkopf und Härtel. (Reprinted by E. J. Bonset 1964.)

49 Weber, E. H. 1834 *De pulsu, resorptione, auditu et tactu. Annotationes anatomicae et physiologicae.* Leipzig, Germany: Koehler.

50 Reid, C. R., Sumpter, D. J. T. & Beekman, M. 2011 Optimisation in a natural system: Argentine ants solve the Towers of Hanoi. *J. Exp. Biol.* **214**, 50–58. (doi:10.1242/jeb.048173)

51 Reynolds, C. W. 1987 Flocks, herds and schools: a distributed behavioral model. *SIGGRAPH Comput. Graph.* **21**, 25–34. (doi:10.1145/37402.37406)

52 Couzin, I. D., Krause, J., James, R., Ruxton, G. D. & Franks, N. R. 2002 Collective memory and spatial sorting in animal groups. *J. Theor. Biol.* **218**, 1–11. (doi:10.1006/jtbi.2002.3065)

53 Chaté, H., Ginelli, F., Grégoire, G., Peruani, F. & Raynaud, F. 2008 Modeling collective motion: variations on the Vicsek model. *Eur. Phys. J. B* **64**, 451–456. (doi:10.1140/epjb/e2008-00275-9)

54 Ginelli, F. & Chaté, H. 2010 Relevance of metric-free interactions in flocking phenomena. *Phys. Rev. Lett.* **105**, 168103. (doi:10.1103/PhysRevLett.105.168103)

55 Bode, N. W. F., Franks, D. W. & Wood, A. J. 2011 Limited interactions in flocks: relating model simulations to empirical data. *J. R. Soc. Interface* **8**, 301–304. (doi:10.1098/rsif.2010.0397)

56 Strombom, D. 2011 Collective motion from local attraction. *J. Theor. Biol.* **283**, 145–151. (doi:10.1016/j.jtbi.2011.05.019)

57 Hemelrijk, C. K. & Hildenbrandt, H. 2011 Some causes of the variable shape of flocks of birds. *PLoS ONE* **6**, e22479. (doi:10.1371/journal.pone.0022479)

58 Kunz, H. & Hemelrijk, C. K. 2012 Simulations of the social organization of large schools of fish whose perception is

obstructed. *Appl. Anim. Behav. Sci.* **138**, 142–151. (doi:10.1016/j.applanim.2012.02.002)

59 Ballerini, M. 2008 Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Anim. Behav.* **76**, 201–215. (doi:10.1016/j.anbehav.2008.02.004)

60 Cavagna, A., Cimarelli, A., Giardina, I., Parisi, G., Santagati, R., Stefanini, F. & Viale, M. 2010 Scale-free correlations in starling flocks. *Proc. Natl Acad. Sci. USA* **107**, 11 865–11 870. (doi:10.1073/pnas.1005766107)

61 Hemelrijk, C. K., Hildenbrandt, H., Reinders, J. & Stamhuis, E. J. 2010 Emergence of oblong school shape: models and empirical data of fish. *Ethology* **116**, 1099–1112. (doi:10.1111/j.1439-0310.2010.01818.x)

62 Katz, Y., Tunstrom, K., Ioannou, C. C., Huepe, C. & Couzin, I. D. 2011 Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl Acad. Sci. USA* **108**, 18 720–18 725. (doi:10.1073/pnas.1107583108)

63 Herbert-Read, J. E., Perna, A., Mann, R. P., Schaerf, T. M., Sumpter, D. J. T. & Ward, A. J. W. 2011 Inferring the rules of interaction of shoaling fish. *Proc. Natl*

Acad. Sci. USA **108**, 18 726–18 731 (doi:10.1073/pnas.1109355108)

64 Mann, R., Freeman, R., Osborne, M., Garnett, R., Armstrong, C., Meade, J., Biro, D., Guilford, T. & Roberts, S. J. 2011 Objectively identifying landmark use and predicting flight trajectories of the homing pigeon using Gaussian processes. *J. R. Soc. Interface* **8**, 210–219. (doi:10.1098/rsif.2010.0301)

65 Eriksson, A., Nilsson Jacobi, M., Nyström, J. & Tunström, K. 2010 Determining interaction rules in animal swarms. *Behav. Ecol.* **21**, 1106–1111. (doi:10.1093/beheco/arq118)

66 Mann, R. P. 2011 Bayesian inference for identifying interaction rules in moving animal groups. *PLoS ONE* **6**, e22827. (doi:10.1371/journal.pone.0022827)

67 Mann, R. P. 2010 Prediction of homing pigeon flight paths using Gaussian processes. PhD thesis, University of Oxford, UK.

68 Weitz, S., Blanco, S., Fournier, R., Gautrais, J., Jost, C. & Theraulaz, G. 2012 Modeling collective animal behavior with a cognitive perspective: a methodological framework. *PLoS ONE* **7**, e38588. (doi:10.1371/journal.pone.0038588)