APPLICATION

swaRmverse: An R package for the comparative analysis of collective motion

Marina Papadopoulou[1,2] | Simon Garnier[3] | Andrew J. King[1]

[1]Biosciences, School of Biosciences, Geography and Physics, Faculty of Science and Engineering, Swansea University, Swansea, UK

[2]Department of Ecological and Biological Sciences, University of Tuscia, Viterbo, Italy

[3]Department of Biological Sciences, New Jersey Institute of Technology, Newark, New Jersey, USA

Correspondence
Marina Papadopoulou
Email: m.papadopoulou.rug@gmail.com

Abstract

1. Collective motion, that is the coordinated spatial and temporal organisation of individuals, is a core element in the study of collective animal behaviour. The self-organised properties of how a group moves influence its various behavioural and ecological processes, such as predator–prey dynamics, social foraging and migration. However, little is known about the inter- and intra-specific variation in collective motion. Despite the significant advancement in high-resolution tracking of multiple individuals within groups, providing collective motion data for animals in the laboratory and the field, a framework to perform quantitative comparisons across species and contexts is lacking.

2. Here, we present the swaRmverse package. Building on two existing R packages, trackdf and swaRm, swaRmverse enables the identification and analysis of collective motion 'events', as presented in Papadopoulou et al. (2023), creating a unit of comparison across datasets. We describe the package's structure and showcase its functionality using existing datasets from several species and simulated trajectories from an agent-based model.

3. From positional time-series data for multiple individuals (x-y-t-id), swaRmverse identifies events of collective motion based on the distribution of polarisation and group speed. For each event, a suite of validated biologically meaningful metrics are calculated, and events are placed into a 'swarm space' through dimensional reduction techniques.

4. Our package provides the first automated pipeline enabling the analysis of data on collective behaviour. The package allows the calculation and use of complex metrics for users without a strong quantitative background and will promote communication and data-sharing across disciplines, standardising the quantification of collective motion across species and promoting comparative investigations.

KEYWORDS
animal behaviour, collective motion, comparative analysis, dimensionality reduction, model validation, R package, trajectory data

1 | INTRODUCTION

The study of collective behaviour is a highly interdisciplinary field of research; biologists, psychologists, physicists, complexity scientists, engineers, and others combine their expertise to understand how individuals behave in groups and how their social interactions lead to complex collective patterns (Couzin & Krause, 2003; Garnier et al., 2007; Herbert-Read, 2016; Vicsek & Zafeiris, 2012). Whether examining fish schools (Georgopoulou et al., 2022; Herbert-Read et al., 2011; Katz et al., 2011), sheep flocks (Ginelli et al., 2015; King et al., 2012), or human crowds (Ettehadieh et al., 2014; King et al., 2015; Moussaid et al., 2009), researchers often analyse data comprising multiple individuals' trajectories collected over time using motion tracking from animal-attached tags or camera images (Dell et al., 2014; Fehlmann & King, 2016; King et al., 2018). Despite increasing data availability, comparisons across datasets are still lacking (Papadopoulou, Fürtbauer, et al., 2023), hindering progress in advancing our understanding of the mechanisms, functions, and evolution of collective behaviour (Biro et al., 2016; Bousquet et al., 2024; Ioannou & Laskowski, 2023; Jolles et al., 2020; King et al., 2018).

The coordinated spatial and temporal organisation of individuals seen in animal groups, whether it be fish schools or primate troops, can be studied as collective motion (Vicsek & Zafeiris, 2012), and understood from the perspective of self-organisation, where group properties emerge from individual rules of motion and interaction (Camazine et al., 2003; Garnier et al., 2007; Hemelrijk & Hildenbrandt, 2011). Collective motion is a common aspect of many collective behaviours in nature, particularly when groups move together to migrate (Voelkl et al., 2015), forage (King et al., 2008, 2015; Mazué et al., 2023) or escape predators (Herbert-Read et al., 2015; King et al., 2012; Papadopoulou et al., 2022). Metrics of collective motion, such as the group order (how aligned its members are) and density (number of individuals per unit area), are often used to quantify the level of synchrony and cohesion within the group (Huth & Wissel, 1994; Inada & Kawachi, 2002), and these simple metrics can provide researchers insight into how group properties relate to group functioning (Giardina, 2008; Sumpter, 2006; Sumpter et al., 2012).

Despite the existence of established metrics that capture the characteristics of collective motion across species (Katz et al., 2011; Lukeman et al., 2010; Pettit et al., 2013; Herbert-Read et al., 2011), their use often relates to a specific field of research. For example, how group cohesion in fish schools or bird flocks changes under predation risk (Herbert-Read et al., 2017; Sankey, Storms, et al., 2021) or how social relationships affect the collective movement of troops of primates (Bracken et al., 2022; Farine et al., 2016) or herds of ungulates (Ozogány et al., 2023; Sankey, O'Bryan, et al., 2021; Torney et al., 2018). We thus lack a standardised framework for quantifying inter- and intra-specific variation in collective motion. In addition, given the diverse background and range of quantitative skills of collective behaviour researchers, analysing trajectories requires substantial effort and time, leading to only a few metrics being used

per study. This further introduces disparities across studies and increases the risks of software errors. Overall, comparative analyses that can unify research practices and generalise findings across systems are hindered.

The animal behaviour community would therefore benefit from an automated pipeline enabling the analysis of data on collective motion. An approach to studying the intra-and inter-specific variation in collective motion across species and contexts has recently been proposed by Papadopoulou, Fürtbauer, et al. (2023). Here, we extend and develop this framework to provide an R package for analysing collective motion datasets. Our package requires positional time-series data for multiple individuals (x-y-t-id), from which it identifies events of collective motion. For each event, a suite of biologically meaningful metrics are calculated. Events are placed into a 'swarm space': a multi-dimensional space in which each dimension captures a single characteristic of collective motion, visualised through dimensionality reduction techniques, allowing a straightforward comparison of events within- and across- groups, species, or contexts. We demonstrate the functionality and applicability of swaRmverse using a dataset of goat herds (*Capra aegagrus hircus*; O'Bryan et al., 2019; Sankey, O'Bryan, et al., 2021), a simulated dataset from an agent-based model of homing pigeons (*Columba livia*; Papadopoulou et al., 2022), and metrics of collective motion that are included in the package, calculated from three previously published datasets of stickleback fish (*Gasterosteus aculeatus*; Georgopoulou et al., 2022), homing pigeons (Papadopoulou et al., 2022; Sankey, Storms, et al., 2021), and chacma baboons (*Papio ursinus*; Bracken et al., 2022) in Papadopoulou, Fürtbauer, et al. (2023). We further present a step-by-step case study, comparing tracks of a herd of goats and a flock of sheep moving in the same area.

2 | INTO THE SWARMVERSE

The swaRmverse package builds on two existing packages that facilitate the loading and processing of spatiotemporal data of collectives: trackdf (Garnier, 2023b) and swaRm (Garnier, 2023a). We use trackdf to standardise the input format of the data into the package pipeline. We use swaRm to process motion data at a lower level, for instance, calculating the heading time-series of an individual, the distance between all individuals, or the bearing angle of each individual to its nearest neighbour at a single time point. The swaRmverse package then combines these into a complete analysis pipeline allowing researchers to quickly investigate characteristics of collective motion, according to the framework developed by Papadopoulou, Fürtbauer, et al. (2023).

The package is available on *CRAN* and can be installed through the install.packages function in R (R Core Team, 2019). Any issues should be reported on the GitHub repository: https://github.com/marinapapa/swaRmverse. Its current version (v0.1.1) requires R version ≥3.5. Instructions on how to prepare a dataset for the package pipeline, calculate metrics of collective motion, and construct 'Swarm spaces' (see below) are included in our vignettes

('https://marinapapa.github.io/swaRmverse/articles/') as 'step1_data_preparation.html', 'step2_data_analysis.html', and 'step3_swarm_space.html', respectively.

2.1 | Package workflow

The complete pipeline of swaRmverse is built around five steps (Figure 1a). The data are first transformed into the standardised format required by the package (Figure 1a.1). This is done by the set_data_format function that requires vectors of positional data (x and y, or longitude and latitude), individual identifiers, and time. A *set* variable is used to specify independent periods of collective motion, for instance, the date of each trajectory (default), or the id of its experimental trial (user-defined). Many specifications about the presence of another grouping level in the data can thus be added through the set_data_format function, for example, if one wants to compare trajectory data for fish schools in different test arenas, or schools of different group sizes. All variables are combined into the *set* column. Then, the add_velocities function calculates the heading and speed of each individual, adds them to the working data-frame (Figure 1a.2), and returns a list of data-frames with each element corresponding to a single *set*. Since headings are inferred from the positional time-series, when an individual is not moving (very low speed), its heading is set to *NA*.

With the main time-series list complete, two functions calculate a series of higher-level metrics. First, time-series of bearing angle and distance to each individual's nearest neighbour are added by the
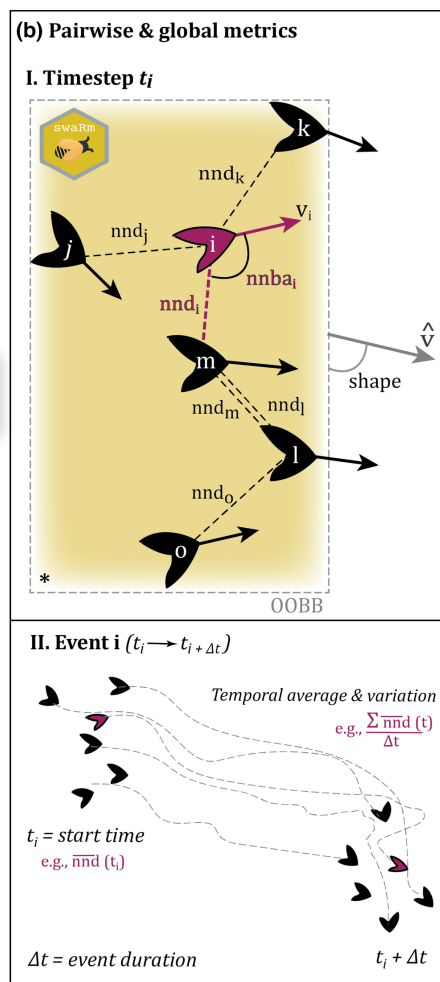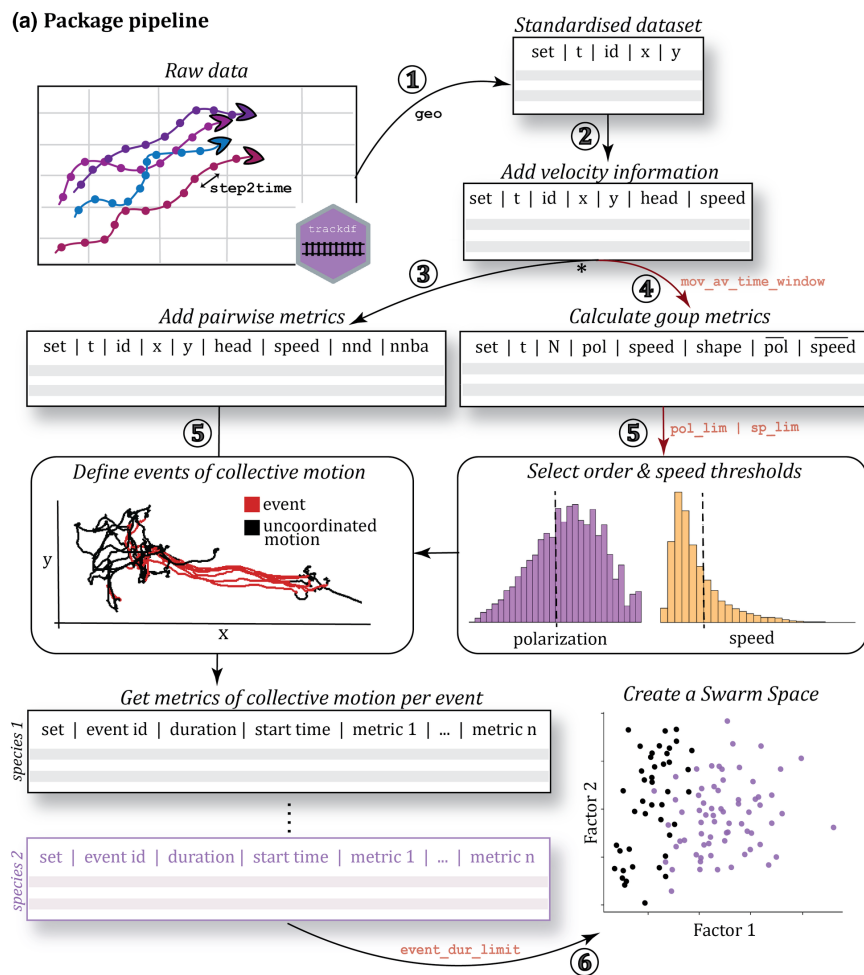

FIGURE 1 Schematic overview of swaRmverse. (a) The main pipeline. Each number links to a major function of the package, as also given in Table 1. Each table represents a data.frame or list of data-frames in the package, and the parameters *step2time*, *geo*, *mov_av_time_window*, *pol_lim*, *sp_lim*, and event_dur_limit the user-defined arguments of each step. Red arrows and parameter names indicate steps that require an informed decision from the user or suggested sensitivity analysis for the chosen values. The $\overline{pol}$ and $\overline{speed}$ columns of the group metrics table are the moving average of polarisation and speed of the group over a given time window. (b.I) Metrics calculated at each time-step, linked to steps 3 and 4 of (a). The nnds indicate the distance to the nearest neighbour of each group member, and nnba the bearing angle between a focal individual i and its nearest neighbour. v_i and $\hat{v}$ indicate the velocity of a focal individual and the group at a given t_i, respectively. The shape index is shown as the angle between the group velocity vector and the longest side of an object-oriented bounding box around all individuals. (b.II) Metrics for each event calculated at step 5 of (a).

pairwise_metrics function (Figure 1a.3,b.I). Then a new data-frame with time-series of the group-level measurements of polarisation, shape, and speed is created by the group_metrics_per_set function (Figure 1a.4,b.I). To remove noise from the group time-series, a time window can be given as argument to calculate the moving average of polarisation and speed over the time-series of each set. Based on these two data-frames, the col_motion_metrics function splits the time-series into events of collective motion and calculates 10 metrics for each event (Figure 1a.5,b.II). The definition of events of collective motion, metrics, and the creation of a swarm space (Figure 1a.6) are explained in detail below. The major functions currently included in the package are summarised in Table 1.

2.2 | Events of collective motion

To enable meaningful comparisons across datasets, a unit of comparison should first be chosen. During the past 20 years, high-resolution tracking data of multiple individuals' positions within groups have been collected through a variety of tools. Laboratory-based research tends to use short observations and image-based tracking designed to gather fine-scale data on variation in specific types of collective motion (Biro et al., 2016; Ioannou & Laskowski, 2023). In contrast, research on wild animals aims to maximise the duration of data collection, using sensor technologies to capture a range of

collective behaviour events (Hughey et al., 2018; King et al., 2018; Westley et al., 2018). Therefore, while the data needed to conduct comparative analyses exist, they are in different forms and often 'hidden' within larger movement datasets.

swaRmverse extracts discrete 'events' of collective motion from the time-series data based on the distribution of polarisation (degree of alignment of group members) and group speed, as developed by Papadopoulou, Fürtbauer, et al. (2023). To remove noise and allow the identification of long continuous periods of coordinated motion, the col_motion_metrics function uses the smoothed time-series of polarisation and speed according to the user-defined moving average time-window of group_metrics_per_set. Based on these distributions, the user can decide the event threshold of each variable: time-periods with lower values in speed or polarisation than the respective thresholds will be discarded, and consecutive time-steps above the thresholds will be labelled as belonging to the same 'event'. A noise threshold can also be selected to merge consecutive events, if there is a small time-window separating them.

The events definition is an important decision to be taken by the user and should be guided by the nature and aim of the analysis. Since it aims to represent the unit of collective motion of a given species, for comparisons within a species, choosing a single threshold across datasets is recommended. However, given that different species will vary in the proportion of time that individuals spend in coordinated motion and in how ordered their

TABLE 1 The major functions comprising the current pipeline of the swaRmverse package, along with the necessary user-defined parameters.

Function	Description	Parameters	Step
Main pipeline			
set_data_format	Standardises the raw data input to the main data-frame with individual positional time-series.	*raw data information*	(1)
add_velocities	Adds the time-series of heading and speed in the dataset.	*geo*	(2)
pairwise_metrics	Adds the time-series of distance and bearing angle of each individual to its nearest neighbour (NN) in a dataset.	*geo*	(3)
group_metrics_per_set	Calculates the time-series of group properties (group size, polarisation, shape, average speed).	*geo, sampling frequency, sliding window*	(4)
col_motion_metrics	Splits the data in events of collective motion and calculates the metrics for each event.	*sampling frequency, speed & polarisation thresholds*	(5)
swarm_space	Splits the data in events of collective motion and calculates the metrics for each event.	type of space ("*pca*" or "*tsne*"), *t-SNE parameters, event duration filter*	(6)
Further functionality			
group_metrics	Calculates the time-series of collective properties of a group.	*geo, sampling frequency*	(3)
nn_metrics	Adds the time-series of distance and bearing angle of each individual to its NN in a group.	*geo*	(4)
expand_pca_swarm_space	Adds new data points in a previously created pca space.	*new event metrics, pca object*	(6)

Note: The *raw data information* refers to the timeseries of positions (with individual identifiers) and details about the type of data, as requested by the track_df package. The *geo* argument captures whether the positional data are geographic coordinates. The *sliding window* is the averaging window for the estimation of the smoothed timeseries of polarisation and group speed. The *new event metrics* refers to a data-frame with collective motion metrics, as returned by the col_motion_metrics function. The exact list of all arguments with their descriptions are given in Table S1.

groups are (Papadopoulou, Fürtbauer, et al., 2023; Papadopoulou, Hildenbrandt, & Hemelrijk, 2023), these thresholds can be adjusted accordingly. Similarly, for pre-processed experimental data focusing only on collective motion, cutting the trajectories of each *set* may not be meaningful; the thresholds can then be set to 0 and each event will be a full *set* as defined in set_data_format. The duration of each event, also exported by the col_motion_metrics function, and the total number of identified events can be used to validate that the threshold selection is meaningful, capturing real periods of collective motion. A balance between the quantity and duration of identified events given varying thresholds should be examined (Figure 2) to ensure that the chosen thresholds provide useful information for answering the user's research question. In addition, through the noise_thresh argument of the col_motion_metrics function, the user can choose to merge events that are separated by a small time window. The start time of each event, included in the exported data-frame, can also be used to visualise the trajectories of each event.

2.3 | Metrics of collective motion

Variations in the way that individuals move as a collective can exist across many levels. At a fine scale, individuals have a given distance and angle to their neighbours at every point in time. At a global level, the group has a shape in space, a given degree of order, as well as the averages and within-group variation in the aforementioned pairwise metrics. At the scale of an event, the group may vary on all these characteristics over time. To capture inter- and intra-specific variation across these levels, at its current state, the package calculates 10 metrics described below.

The order of a group, how aligned group members are to each other, and its variation over time, are captured through the average and the standard deviation of polarisation over time during each event (mean_pol and sd_pol), respectively (Figure 1b). The cohesion of the group is captured by three metrics. First, based on the distance of each group member to its nearest neighbour at a given time-point, we calculate the time-series of average nearest neighbour distance (NND) during an event (Figure 1b). Then, we calculate the overall average NND on an event (mean_mean_nnd) and its variation over time (sd_mean_nnd). Second, to capture within-group variation in NND, we calculate its standard deviation at each time-step, and its average during an event (mean_sd_nnd). We further measure how much a group changes its speed during an event by calculating the coefficient of variation in average speed over time (cv_speed), to also allow comparisons across species with different locomotion styles and speed magnitude.

Characteristics of a group's spatiotemporal dynamics are also measured through metrics of shape and internal structure. We use an object-oriented bounding box (OOBB) approach (as in Hemelrijk & Hildenbrandt, 2011) to capture whether a group is more wide than long or more long than wide. Based on the OOBB that contains all group members, we calculate the angle between its longest side and the average heading of the group (Figure 1b.I). Thus our shape index varies from 0 rads (for a perfectly oblong group) to pi/2 rads (for a perfectly wide group). We measure the average and the standard deviation of this index over time during each event (mean_shape and sd_shape). The relative position of individuals to
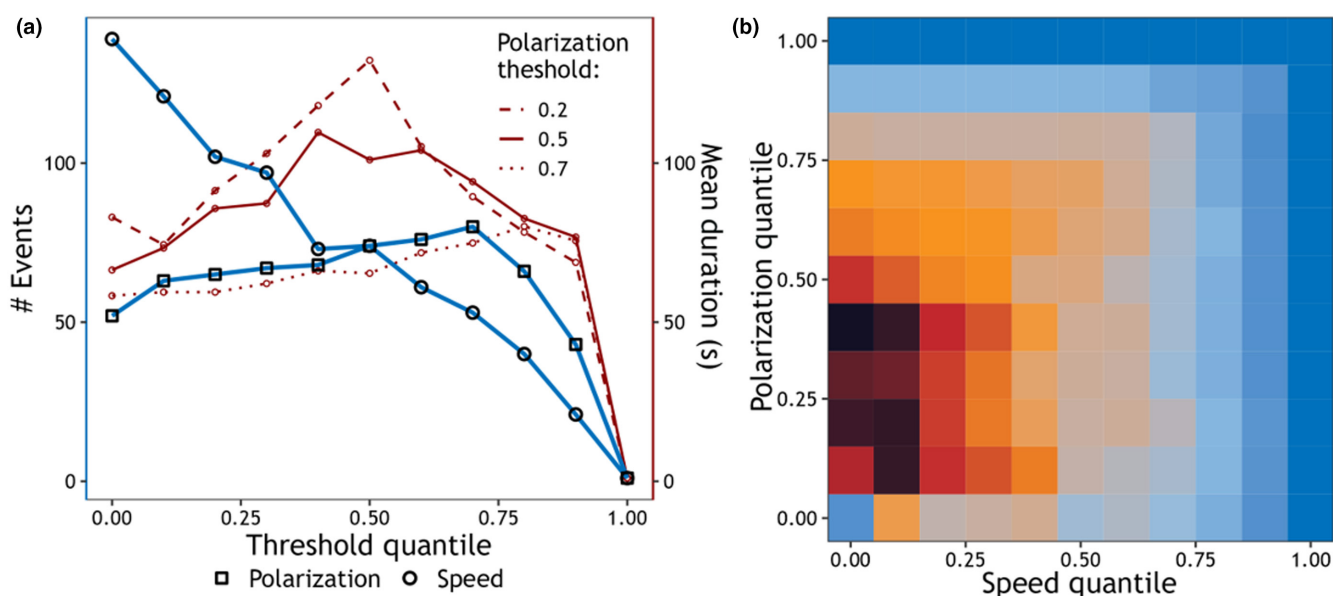

FIGURE 2 (a) Sensitivity analysis of the number (y-left) and average duration (y-right) of events of collective motion over varying thresholds for average speed and polarisation in a herd of goats (O'Bryan et al., 2019; Sankey, O'Bryan, et al., 2021). The solid blue lines represent varying quantile for speed (circles) and polarisation (squares), while the other quantile is set to 0.5. (b) Heatmap of the number of events identified for each combination of speed and polarisation quantile. In Papadopoulou, Fürtbauer, et al. (2023), the 0.5 quantile was used as a threshold for this dataset.

one another is measured through the nearest neighbour bearing angle: the angle between the heading of each focal individual and the line between its position and the position of its nearest neighbour (Figure 1b.I). We calculate the group average of this angle from all group members and the average over time (mean_mean_bangl). Finally, by calculating the frontness (Papadopoulou, Fürtbauer, et al., 2023) between each individual and its nearest neighbour (the absolute value of the bearing angle divided by 180°, that is how in front or behind an individual is positioned relative to its NN), we calculate the standard deviation over all group members at a given time. Through the average of this metric through time (mean_sd_front), we capture how queue-like a group is, adding structure to the more approximate measurement of shape.

2.4 | The swarm space

A user may take the metrics described above and export them for their own bespoke analyses (see further functionality, below). The end-point of our package pipeline is the investigation of the inter- and intra-specific variation of one or more datasets. To achieve that, the swarm_space function allows for comparisons across quantified events of collective motion using dimensionality reduction techniques (Ayesha et al., 2020). The package currently supports principal components analysis (PCA) (Hotelling, 1933 using the stats package) and t-distributed stochastic neighbour embedding (t-SNE) (using the Rtsne package; Krijthe, 2015; van der Maaten & Hinton, 2008). These two methodologies can work complementary to one another, with the first reflecting the global structure in the data and the latter the between-events similarities (Ayesha et al., 2020). However, given the events data-frame, the user can choose other analyses depending on the specifics of the aimed comparison. For instance, if the objective is to compare events across many species from different taxonomic clades, one may consider using Phylogenetic Factor Analysis (PFA, Hassler et al., 2022). The main factors identified from the dimensionality reduction can then be examined to see which variables explain most variation in the data and load onto the different axes. For a more detailed interpretation and discussion of these methods for comparison across species, see Papadopoulou, Fürtbauer, et al. (2023).

The PCA option further allows to project new data on an existing space (using the prcomp function of stats). This is done by the expand_swarm_space function that takes a previously created PCA object and metrics of collective motion of a new dataset. A PCA object of the four empirical datasets of Figure 4a and the PC1-3 coordinates of each event from Papadopoulou, Fürtbauer, et al. (2023) are included in the package as multi_species_pca and multi_species_pca_data, respectively. With more datasets being analysed with our package, this existing multi-dimensional space can keep expanding (see Section 2.5 for an example). Finally, a limit for the minimum duration that an event should have to be included in the space can also be defined through the event_dur_limit argument of the swarm_space function; we recommend removing events of very short duration given that they may not

be representative of the group's collective motion. In the original analysis of Papadopoulou, Fürtbauer, et al. (2023), events with a duration of less than 15 s were discarded.

2.5 | Case study: Sheep flocks versus goat herds

We follow the steps mentioned above (Table 1; Figure 1, steps 1–6) to compare the collective motion of sheep ($N = 10$) and goats ($N = 15$) in Namibia, moving around the same area in two consecutive years (data provided by Lisa O'Bryan, for details about the area and method of data collection see O'Bryan et al. (2019)). Given that the two species also have the same mode of locomotion, we can perform meaningful comparisons of their collective motion. Any differences between the two should reflect differences in the social interactions between group members or different reactions to the same habitat. The two datasets comprise individual GPS trajectories in similar formats: a data-frame with several columns, including the ones for time (in date-time format), individual identifiers, longitude, and latitude (geo=TRUE). To start, we input these columns as vectors in the set_data_format (step 1), indicating also the time-zone in which the data were collected (using the OlsonNames() format from R's timezones). For working with different data types (e.g. camera tracking in the lab), see our vignettes and the trackdf package (Garnier, 2023b). The function returns the tracks of each group in a standardised format, and assigns a 'set' column (categories that may split the tracks) that here coincide with the date of data collection (Figure 3a). We then calculate the timeseries of heading and speed, creating a list of data-frames for each set using the add_velocities function (step 2).

From this list, we can calculate pairwise metrics (with the pairwise_metrics function): adding the distance, id and bearing angle of each individual's nearest neighbour to the timeseries data-frame (step 3). By investigating the distribution of these variables (e.g. Figure 3b), we noticed a few errors in the GPS locations of the raw sheep data. We thus filtered the data to exclude any position that had nearest neighbour distance larger than 50 meters. The user should consider whether data cleaning or trajectory smoothing practices are necessary for their dataset or research question. Based on the clean data, we calculate the group-level timeseries of polarisation, speed and shape (e.g. Figure 3c), using the group_metrics_per_set function, with a chosen averaging sliding window (here 30 s, step 4). We can then use the col_motion_metrics function to split the tracks in events of collective motion and calculate all our metrics (step 5). We used the 0.5 and 0.75 quantile of the distribution of speed and polarisation, respectively, per species to define our events (but see Section 2.2 and Figure 2 for information in selecting these thresholds).

After filtering out events with duration less than 15 s (Papadopoulou, Fürtbauer, et al., 2023), we merged the data-frames of each species and we performed the dimensionality reduction with the swarm_space function (step 6), using the 'pca' option, to compare the events of the two species (Figure 3c). The function returns a list with three elements: the 'swarm_space' data-frame (with PC1-3
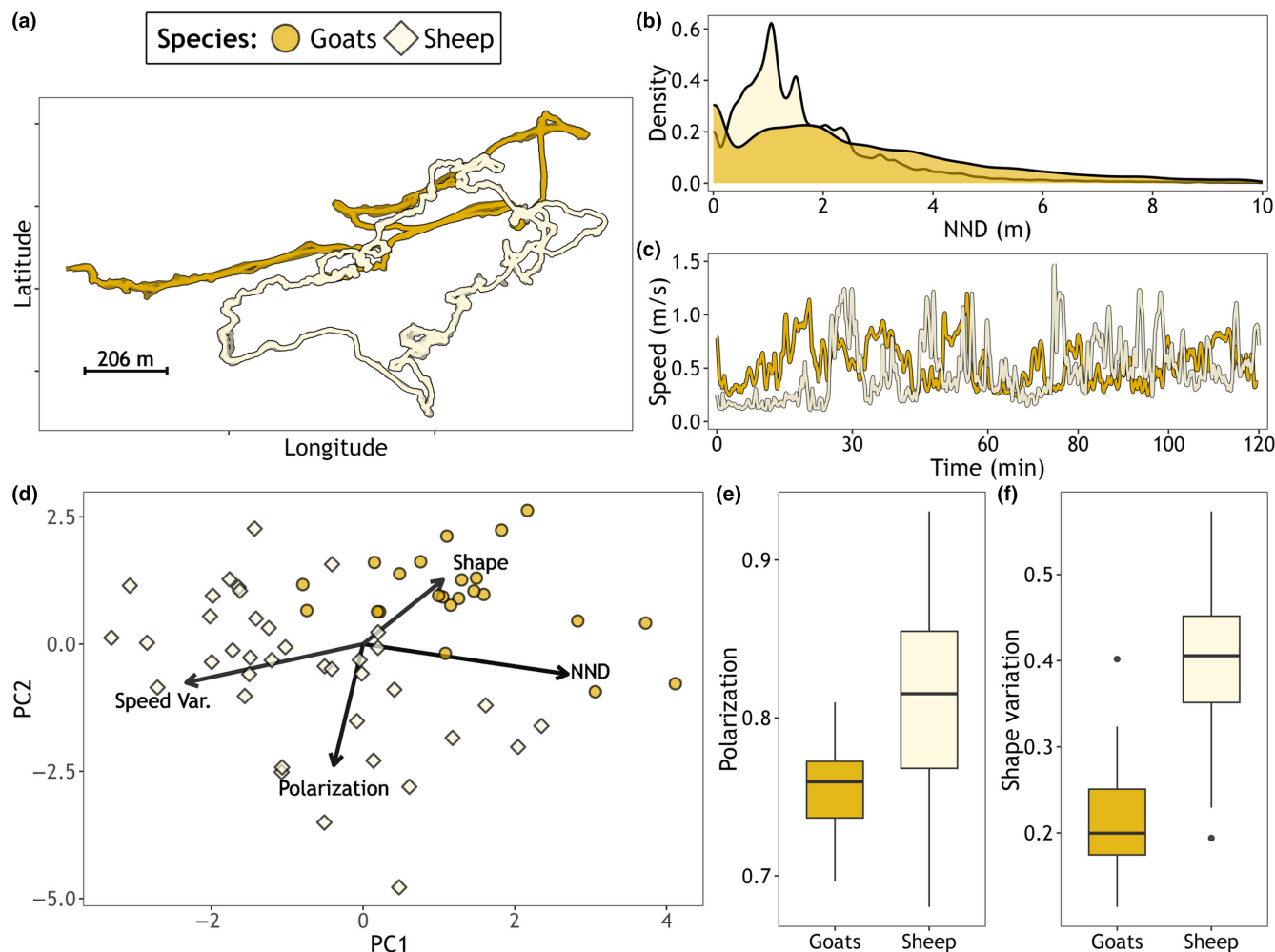
FIGURE 3 Case study: Goats versus sheep. (a) The trajectories of each group. (b) The distributions of nearest neighbour distance (NND). (c) The timeseries of average speed of each group, smoothed using a 30 s time window. (d) The PCA comparing the identified events of collective motion, with the loadings of some important variables (mean_shape, mean_mean_nnd, mean_pol, cv_speed). (e) Polarisation and (f) variation in shape during the events of each species.

values per event), a 'ref' data-frame that holds the additional information of each event (e.g. starting time, duration), and the PCA object. By plotting the *swarm_space* data (Figure 3d), we see that the two species differ in several characteristics such as average polarisation and temporal variation in their group shape (Figure 3e,f). The first three components of the PCA explain approximately 65% of variance, something the user should investigate using the summary function on the PCA object. Finally, we can expand our previous swarm space (multi_species_pca) to include the sheep dataset (Figure 4a), using the expand_swarm_space function. All code to reproduce this case study, including the plotting functions (not available in the package), is on our online repository: https://github.com/marinapapa/Papadopoul ou_et_al_2024_swaRmverse (Papadopoulou, 2024).

2.6 | Further functionality

Apart from the analysis of empirical data, the swarm space approach might be particularly useful for comparing simulated

data from agent-based models that are often used in collective behaviour research, to validate whether the model captures the characteristics of the system it aims to study (supporting pattern-oriented modelling, Grimm et al., 2005). Here, we created another example of a new swarm space (Figure 4b) that combines the previously published events of collective motion Papadopoulou, Fürtbauer, et al. (2023), also included in the package (multi_species_metrics), and events from analysing simulated trajectories from the HoPE model, an agent-based model calibrated to resemble flocks of homing pigeons through distributions of speed, nearest neighbour distance and relative position of neighbours (Papadopoulou et al., 2022). Each species covers a different area of the PCA space, with the simulated pigeon flocks overlapping with the real ones (Figure 4b).

The complete pipeline of the package can be accessed through the col_motion_metrics_from_raw function (steps 2–5, used here for the analysis of the simulated data), when only the overall metrics across sets and events are returned, given the standardised input data-frame. Outside of the end goal of the swarm space and
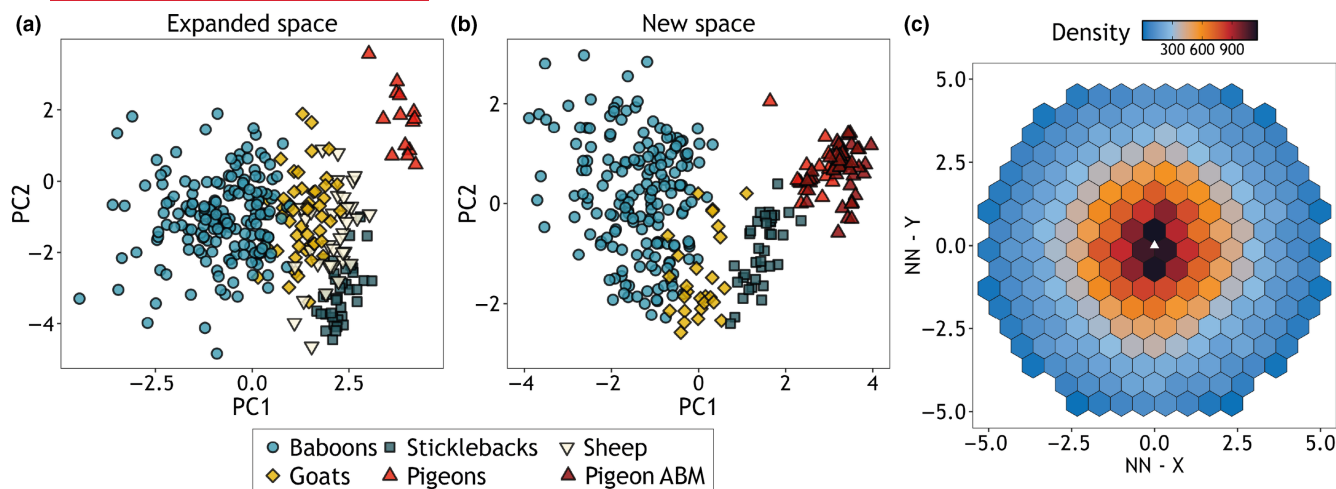
FIGURE 4 (a) A PCA space comparing events of collective motion in five empirical datasets of chacma baboons (Bracken et al., 2022), goats (O'Bryan et al., 2019; Sankey, O'Bryan, et al., 2021), stickleback fish (Georgopoulou et al., 2022), homing pigeons (Sankey, Storms, et al., 2021) and sheep (data by Lisa O'Bryan). The PCA object is the one previously published by (Papadopoulou, Fürtbauer, et al., 2023), with the new data from our case study added through the expand_swarm_space function. (b) A new PCA space comparing the previously published events of collective motion, and a simulated dataset from an agent-based model of pigeons (Papadopoulou et al., 2022). (c) The relative position (nnx, nny) of the nearest neighbour (NN) in the reference frame of all focal individuals in the goats dataset (distance measured in meters).

structure of our pipeline, our package can help with the analysis of spatiotemporal datasets. For instance, the lower level function nn_metrics calculates for each time-step in a data-frame the bearing angle and distance to a nearest neighbour, while the co-ordinates of this neighbour in the reference frame of the focal individual can also be added through the add_coords argument or by the standalone function add_rel_pos_coords (Figure 4c). Similarly, the group_metrics function calculates the time-series of average speed, polarisation, and shape index of a group. A more experienced user can thus combine functionalities and create new metrics to visualise the collective motion characteristics of their system.

These metrics (e.g. NND, group polarisation), apart from their significance in the study of self-organised social systems, are valuable to our understanding of animal movement ecology. For instance, one can focus their analysis on the spatial distribution of the events of collective motion identified by the package or how these metrics vary across the range of the group. Such investigation will provide insights on the way animals interact with their ecosystem and add the dimension of group behaviour in the analysis of GPS and accelerometery data of individuals (Fehlmann et al., 2017). In experiments in the laboratory or the management of livestock, these metrics can also be used to automatically flag conditions related to animal health and welfare (Demandt et al., 2018; Matthews et al., 2017).

3 | CONCLUSIONS

Both the in-between steps and the ending point of our pipeline (the produced swarm space) are valuable to future studies. The swarm

space can act as a first step in the analysis of collective behaviour data, hinting in which direction the analysis should continue, disentangling the source of variation within or across datasets. Our package can also facilitate comparisons between models and empirical data, a crucial and challenging step in the validation of results and theoretical conclusions in self-organised systems (Crooks et al., 2008). These comparisons are usually focused on a couple of metrics, without others being investigated or with clear criteria in selecting them. At the same time, comparisons across many collective characteristics can be challenging (Hildenbrandt et al., 2010; Papadopoulou, Fürtbauer, et al., 2023). The new components of a swarm space or the metrics that capture more variation in a given dataset can then be selected instead, as the key aspects for comparisons between models and data. This can further facilitate bio-inspired applications in swarm-robotics, with the emergent collective characteristics of a given algorithm being clearly linked to a given function (Dorigo et al., 2021).

Apart from the insights of this package for collective behaviour experts, this work is a step towards the standardisation of analysis of collective motion across systems. We hope that with the involvement of the collective behaviour community, the package will expand to include more metrics of collective motion such as path characteristics (Roberts et al., 2004), pairwise information for neighbours of higher order and rate of change of their social network (Papadopoulou, Hildenbrandt, & Hemelrijk, 2023), capturing more variation that may exist across systems. Through the ease of calculating these metrics with any existing dataset, irrespective of the main aim of the study that collected it, we can work towards a mapping of characteristics of collective motion and start building bridges across disciplines that are interested in understanding

self-organised social systems, while also supporting the further use of existing open datasets.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14460.

DATA AVAILABILITY STATEMENT

The package swaRmverse is available on the Comprehensive R Archive Network (CRAN; https://cran.r-project.org/web/packages/swaRmverse/index.html). Raw data used as examples are available via: https://github.com/swarm-lab/goatCollectiveDecision for the goat herd data (O'Bryan et al., 2019; Sankey, O'Bryan, et al., 2021) and via https://doi.org/10.5281/zenodo.4993109 for the simulated pigeon flock data (Papadopoulou et al., 2022). Further data used and all the code to reproduce our analyses are available via https://doi.org/10.5281/zenodo.14019983 (Papadopoulou, 2024).

ORCID

Marina Papadopoulou https://orcid.org/0000-0002-6478-8365
Simon Garnier https://orcid.org/0000-0002-3886-3974
Andrew J. King https://orcid.org/0000-0002-6870-9767

REFERENCES

Ayesha, S., Hanif, M. K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, *59*, 44–58.

Biro, D., Sasaki, T., & Portugal, S. J. (2016). Bringing a time-depth perspective to collective animal behaviour. *Trends in Ecology & Evolution*, *31*, 550–562.

Bousquet, C. A. H., Sueur, C., King, A. J., & O'Bryan, L. R. (2024). Individual and ecological heterogeneity promote complex communication in social vertebrate group decisions. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *379*, 20230204. https://doi.org/10.1098/rstb.2023.0204

Bracken, A. M., Christensen, C., O'Riain, M. J., Fürtbauer, I., & King, A. J. (2022). Flexible group cohesion and coordination, but robust leader-follower roles, in a wild social primate using urban space. *Proceedings of the Royal Society B: Biological Sciences*, *289*, 20212141.

Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2003). *Self-organization in biologican systems*. Princeton University Press.

Couzin, I. D., & Krause, J. (2003). Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, *32*, 1–75.

Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, *32*, 417–430. https://doi.org/10.1016/j.compenvurbsys.2008.09.004

Dell, A. I., Bender, J. A., Branson, K., Couzin, I. D., de Polavieja, G. G., Noldus, L. P. J. J., Pérez-Escudero, A., Perona, P., Straw, A. D., Wikelski, M., & Brose, U. (2014). Automated image-based tracking and its application in ecology. *Trends in Ecology & Evolution*, *29*, 417–428.

Demandt, N., Saus, B., Kurvers, R. H., Krause, J., Kurtz, J., & Scharsack, J. P. (2018). Parasite-infected sticklebacks increase the risk-taking behaviour of uninfected group members. *Proceedings of the Royal Society B*, *285*, 20180956.

Dorigo, M., Theraulaz, G., & Trianni, V. (2021). Swarm robotics: Past, present, and future. *Proceedings of the IEEE*, *109*, 1152–1165.

Ettehadieh, D., Farooq, B., & Saunier, N. (2014). Automated pedestrian data-collection and flow analysis in public spaces. *Transportation Research Procedia*, *2*, 207–212.

Farine, D. R., Strandburg-Peshkin, A., Berger-Wolf, T., Ziebart, B., Brugere, I., Li, J., & Crofoot, M. C. (2016). Both nearest Neighbours and long-term affiliates predict individual locations during collective movement in wild baboons. *Scientific Reports*, *6*, 27704.

Fehlmann, G., & King, A. J. (2016). Bio-logging. *Current Biology: CB*, *26*, R830–R831.

Fehlmann, G., O'Riain, M. J., Hopkins, P. W., O'Sullivan, J., Holton, M. D., Shepard, E. L., & King, A. J. (2017). Identification of behaviours from accelerometer data in a wild social primate. *Animal Biotelemetry*, *5*, 1–11.

Garnier, S. (2023a). *swaRm: Processing collective movement data*. R package version 0.6.0. https://github.com/swarm-lab/swaRm

Garnier, S. (2023b). *trackdf: Data frame class for tracking data*. https://github.com/swarm-lab/trackdf

Garnier, S., Gautrais, J., & Theraulaz, G. (2007). The biological principles of swarm intelligence. *Swarm Intelligence*, *1*, 3–31. https://doi.org/10.1007/s11721-007-0004-y

Georgopoulou, D. G., King, A. J., Brown, R. M., & Fürtbauer, I. (2022). Emergence and repeatability of leadership and coordinated motion in fish shoals. *Behavioral Ecology*, *33*, 47–54.

Giardina, I. (2008). Collective behavior in animal groups: Theoretical models and empirical studies. *HFSP Journal*, *2*, 205–219.

Ginelli, F., Peruani, F., Pillot, M.-H., Chaté, H., Theraulaz, G., & Bon, R. (2015). Intermittent collective dynamics emerge from conflicting imperatives in sheep herds. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 12729–12734. https://doi.org/10.1073/pnas.1503749112

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H. H., Weiner, J., Wiegand, T., & DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, *310*, 987–991.

Hassler, G. W., Gallone, B., Aristide, L., Allen, W. L., Tolkoff, M. R., Holbrook, A. J., Baele, G., Lemey, P., & Suchard, M. A. (2022). Principled, practical, flexible, fast: A new approach to phylogenetic factor analysis. *Methods in Ecology and Evolution*, 13, 2181–2197. https://doi.org/10.1111/2041-210X.13920

Hemelrijk, C. K., & Hildenbrandt, H. (2011). Some causes of the variable shape of flocks of birds. *PLoS One*, 6, e22479.

Herbert-Read, J. E. (2016). Understanding how animal groups achieve coordinated movement. *Journal of Experimental Biology*, 219, 2971–2983.

Herbert-Read, J. E., Buhl, J., Hu, F., Ward, A. J. W., & Sumpter, D. J. (2015). Initiation and spread of escape waves within animal groups. *Royal Society Open Science*, 2, 140355.

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. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 18726–18731.

Herbert-Read, J. E., Rosén, E., Szorkovszky, A., Ioannou, C. C., Rogell, B., Perna, A., Ramnarine, I. W., Kotrschal, A., Kolm, N., Krause, J., & Sumpter, D. J. (2017). How predation shapes the social interaction rules of shoaling fish. *Proceedings of the Royal Society B: Biological Sciences*, 284, 20171126.

Hildenbrandt, H., Carere, C., & Hemelrijk, C. K. (2010). Self-organized aerial displays of thousands of starlings: A model. *Behavioral Ecology*, 21, 1349–1359.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441.

Hughey, L. F., Hein, A. M., Strandburg-Peshkin, A., & Jensen, F. H. (2018). Challenges and solutions for studying collective animal behaviour in the wild. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 373, 20170005. https://doi.org/10.1098/rstb.2017.0005

Huth, A., & Wissel, C. (1994). The simulation of fish schools in comparison with experimental data. *Ecological Modelling*, 75–76, 135–146.

Inada, Y., & Kawachi, K. (2002). Order and flexibility in the motion of fish schools. *Journal of Theoretical Biology*, 214, 371–387.

Ioannou, C. C., & Laskowski, K. L. (2023). A multi-scale review of the dynamics of collective behaviour: From rapid responses to ontogeny and evolution. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 378, 20220059. https://doi.org/10.1098/rstb.2022.0059

Jolles, J. W., King, A. J., & Killen, S. S. (2020). The role of individual heterogeneity in collective animal behaviour. *Trends in Ecology & Evolution*, 35, 278–291.

Katz, Y., Tunstrom, K., Ioannou, C. C., Huepe, C., & Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 18720–18725.

King, A. J., Douglas, C. M. S., Huchard, E., Isaac, N. J. B., & Cowlishaw, G. (2008). Dominance and affiliation mediate despotism in a social primate. *Current Biology*, 18, 1833–1838.

King, A. J., Fehlmann, G., Biro, D., Ward, A. J., & Fürtbauer, I. (2018). Rewilding collective behaviour: An ecological perspective. *Trends in Ecology & Evolution*, 33, 347–357.

King, A. J., Myatt, J. P., Fürtbauer, I., Oesch, N., Dunbar, R. I. M., Sumner, S., Usherwood, J. R., Hailes, S., & Brown, M. R. (2015). Social density processes regulate the functioning and performance of foraging human teams. *Scientific Reports*, 5, 18260.

King, A. J., Wilson, A. M., Wilshin, S. D., Lowe, J., Haddadi, H., Hailes, S., & Morton, A. J. (2012). Selfish-herd behaviour of sheep under threat. *Current Biology*, 22, R561–R562.

Krijthe, J. H. (2015). *Rtsne: T-distributed stochastic neighbor embedding using a Barnes-Hut implementation*. R package version 0.17. https://github.com/jkrijthe/Rtsne

Lukeman, R., Li, Y.-X., & Edelstein-Keshet, L. (2010). Inferring individual rules from collective behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 12576–12580. https://doi.org/10.1073/pnas.1001763107

Matthews, S. G., Miller, A. L., Plötz, T., & Kyriazakis, I. (2017). Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. *Scientific Reports*, 7, 17582.

Mazué, G. P. F., Adams, M. W. D., Seebacher, F., & Ward, A. J. W. (2023). Changes in movement characteristics in response to private and social information acquisition of socially foraging fish. *Animal Behaviour*, 201, 73–86.

Moussaid, M., Helbing, D., Garnier, S., Johansson, A., Combe, M., & Theraulaz, G. (2009). Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences*, 276, 2755–2762. https://doi.org/10.1098/rspb.2009.0405

O'Bryan, L. R., Abaid, N., Nakayama, S., Dey, T., King, A. J., Cowlishaw, G., Rubenstein, D. I., & Garnier, S. (2019). Contact calls facilitate group contraction in free-ranging goats (*Capra aegagrus hircus*). *Frontiers in Ecology and Evolution*, 7. https://doi.org/10.3389/fevo.2019.00073/full

Ozogány, K., Kerekes, V., Fülöp, A., Barta, Z., & Nagy, M. (2023). Fine-scale collective movements reveal present, past and future dynamics of a multilevel society in Przewalski's horses. *Nature Communications*, 14, 5096.

Papadopoulou, M. (2024). Data & Code for: 'swaRmverse: An R package for the comparative analysis of collective motion' (v1.0.0). *Zenodo*. https://doi.org/10.5281/zenodo.14019983

Papadopoulou, M., Fürtbauer, I., O'Bryan, L. R., Garnier, S., Georgopoulou, D. G., Bracken, A. M., Christensen, C., & King, A. J. (2023). Dynamics of collective motion across time and species. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 378, 20220068. https://doi.org/10.1098/rstb.2022.0068

Papadopoulou, M., Hildenbrandt, H., & Hemelrijk, C. K. (2023). Diffusion during collective turns in bird flocks under predation. *Frontiers in Ecology and Evolution*, 11. https://doi.org/10.3389/fevo.2023.1198248

Papadopoulou, M., Hildenbrandt, H., Sankey, D. W. E., Portugal, S. J., & Hemelrijk, C. K. (2022). Self-organization of collective escape in pigeon flocks. *PLoS Computational Biology*, 18, e1009772. https://doi.org/10.1371/journal.pcbi.1009772

Pettit, B., Perna, A., Biro, D., & Sumpter, D. J. (2013). Interaction rules underlying group decisions in homing pigeons. *Journal of the Royal Society Interface*, 10, 20130529.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roberts, S., Guilford, T., Rezek, I., & Biro, D. (2004). Positional entropy during pigeon homing I: Application of Bayesian latent state modelling. *Journal of Theoretical Biology*, 227, 39–50.

Sankey, D. W., Storms, R. F., Musters, R. J., Russell, T. W., Hemelrijk, C. K., & Portugal, S. J. (2021). Absence of 'selfish herd' dynamics in bird flocks under threat. *Current Biology*, 34, 3192–3198.e7.

Sankey, D. W. E., O'Bryan, L. R., Garnier, S., Cowlishaw, G., Hopkins, P., Holton, M., Fürtbauer, I., & King, A. J. (2021). Consensus of travel direction is achieved by simple copying, not voting, in free-ranging goats. *Royal Society Open Science*, 8, 201128.

Sumpter, D. J. (2006). The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 361, 5–22.

Sumpter, D. J., Mann, R. P., & Perna, A. (2012). The modelling cycle for collective animal behaviour. *Interface Focus*, 2, 764–773.

Torney, C. J., Lamont, M., Debell, L., Angohiatok, R. J., Leclerc, L.-M., & Berdahl, A. M. (2018). Inferring the rules of social interaction in migrating caribou. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 373, 20170385. https://doi.org/10.1098/rstb.2017.0385

van der Maaten, L., & Hinton, G. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.

Vicsek, T., & Zafeiris, A. (2012). Collective motion. *Physics Reports*, *517*, 71–140.

Voelkl, B., Portugal, S. J., Unsöld, M., Usherwood, J. R., Wilson, A. M., & Fritz, J. (2015). Matching times of leading and following suggest cooperation through direct reciprocity during v-formation flight in ibis. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 2115–2120.

Westley, P. A., Berdahl, A. M., Torney, C. J., & Biro, D. (2018). Collective movement in ecology: From emerging technologies to conservation and management. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 373*, 20170004. https://doi.org/10.1098/rstb.2017.0004

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1: The main arguments of the major functions comprising the current pipeline of the swaRmverse package. The option to parallelize the calculations across sets is given.