Behavioral Ecology Advance Access published October 3, 2016

Behavioral Ecology

The official journal of the ISBE International Society for Behavioral Ecology

Behavioral Ecology (2016), 00(00), 1–11. doi:10.1093/beheco/arw145

Original Article

Direction matching for sparse movement data sets: determining interaction rules in social groups

Tyler R. Bonnell,[a,b] **S. Peter Henzi,**[a,b] **and Louise Barrett**[a,b]

[a]Department of Psychology, University of Lethbridge, C866 University Hall, 4401 University Drive, Alberta T1K 3M4, Canada and [b]Applied Behavioural Ecology and Ecosystems Research Unit, University of South Africa, Private Bag X6, Florida 1710, Republic of South Africa

Received 5 February 2016; revised 12 July 2016; accepted 18 July 2016.

It is generally assumed that high-resolution movement data are needed to extract meaningful decision-making patterns of animals on the move. Here we propose a modified version of force matching (referred to here as direction matching), whereby sparse movement data (i.e., collected over minutes instead of seconds) can be used to test hypothesized forces acting on a focal animal based on their ability to explain observed movement. We first test the direction matching approach using simulated data from an agent-based model, and then go on to apply it to a sparse movement data set collected on a troop of baboons in the DeHoop Nature Reserve, South Africa. We use the baboon data set to test the hypothesis that an individual's motion is influenced by the group as a whole or, alternatively, whether it is influenced by the location of specific individuals within the group. Our data provide support for both hypotheses, with stronger support for the latter. The focal animal showed consistent patterns of movement toward particular individuals when distance from these individuals increased beyond 5.6 m. Although the focal animal was also sensitive to the group movement on those occasions when the group as a whole was highly clustered, these conditions of isolation occurred infrequently. We suggest that specific social interactions may thus drive overall group cohesion. The results of the direction matching approach suggest that relatively sparse data, with low technical and economic costs, can be used to test between hypotheses on the factors driving movement decisions.

Key words: **baboon, behavior, De Hoop Nature Reserve, force matching, optimization, social groups, sparse movement data.**

INTRODUCTION

Fine-grained analyses of the movement strategies of mobile animals have benefited enormously from technical advances in global positioning systems (GPS). Increased miniaturization, higher frequency capture rates, and longevity of battery life all permit the collection of rich data sets from animal-mounted GPS units (Nathan et al. 2008; Cagnacci et al. 2010). In association with the development of new analytical techniques (Laube et al. 2005; Gurarie et al. 2009; Dalziel et al. 2015; De Groeve et al. 2015), such data are beginning to provide answers to long-standing questions in movement ecology and are driving new research programs, particularly with respect to social animals (Lukeman et al. 2010; Hebblewhite and Haydon 2010; Krause et al. 2013; Kays et al. 2015; Strandburg-Peshkin et al. 2015).

There are, however, some limitations associated with the use of animal-attached GPS devices, not all of which can be addressed through technical developments. Although battery life may improve, allowing more than short high-resolution snapshots of activity (Strandburg-Peshkin et al. 2015), equipment failure will persist. Moreover, the need to attach devices to the animals can raise ethical, logistical, and technical issues (Handcock et al. 2009). There are also economic costs that may limit access to some researchers or require sampling of only a very few individuals. Even in cases where one can capture all members of a social group, repeated immobilization to affix and retrieve collars and data may limit researchers' ability to rely on these techniques alone.

Many social animals, most notably primates, can be habituated and followed sufficiently closely to allow human observers to mimic automatic spatial data collection with handheld GPS dataloggers (Sugiura et al. 2011; Aureli et al. 2012; Heesen et al. 2015). This method reduces costs and may lift other logistical constraints, but the resolution of the data (i.e., number of records per unit time) is far lower than that deliverable by fully automated, animal-mounted techniques. Nevertheless, if such data can be shown to produce reliable spatiotemporal movement patterns, then low-resolution

---

Address correspondence to T.R. Bonnell. E-mail: Tyler.bonnell@uleth.ca.

approaches may offer a sustainable, flexible, and reliable alternative approach to automated methods.

Here, we address the extent to which the individual actions that underpin observed collective motion can be detected as the temporal resolution of data is lowered. To do so, we implement a modified version of Eriksson et al.'s (2010) force matching method, which we refer to as "direction matching." This approach sets behavioral rules that determine motion and then fits parameters to these rules that minimize the deviation between observed and predicted movements. These rules are used both to describe those elements of the environment to which an individual is sensitive (i.e., will respond to by moving) and the kinds of behavioral response elicited. Here, the "force" in direction matching can be considered to be an estimate of the magnitude of this sensitivity.

To test how the temporal resolution of data affects the reliability of movement trajectories, we use agent-based modeling to specify the rules governing individual movement in a simulated group of agents. We specify these rules based on the empirical findings described by Strandburg-Peshkin et al. (2015), obtained using high-resolution data from animal-attached GPS units. We record the resultant trajectories of our simulated agents based on similar high-frequency sampling and then subsample these trajectories to generate data sets of differing temporal resolution. We then apply direction matching to each of these and assess the extent to which the results obtained are able to identify the rule that we had built into the agents. In this way, we can determine the extent to which the direction matching approach is able to: 1) identify the true conditions under which a given behavior occurs and 2) accurately identify the correct behavior as temporal resolution declines.

As a second step, we apply this approach to real-world movement data collected from a wild troop of baboons (*Papio hamadryas ursinus*). We offer a "proof of concept" illustration of the way in which our method enables researchers to test between different hypotheses explaining observed movement. Current theories of group movement are based on both leadership and consensus formation within the group (Conradt and Roper 2005; King et al. 2008). In baboons, for example, it has been shown that individuals are sensitive both to the number of initiators and to their directional agreement (DA; Strandburg-Peshkin et al. 2015). There is also evidence to suggest that individuals are more likely to follow those with whom they share close social affiliations (King et al. 2011). These potential rules of thumb, for example, follow close social affiliates or follow the majority of the group, are not necessarily mutually exclusive, which presents a real challenge when attempting to identify or differentiate between the possible mechanisms behind observed motions. For example, in the case of 2 clusters of initiators within a group, a larger group of initiators may be more likely to contain close social affiliates than a smaller group of initiators. This might lead to the observation that, on average, individuals move more toward the majority when, in fact, patterns of affiliation between individuals might be responsible. Using data from a focal individual as a case study, we therefore test the extent to which the movement of the focal animal is 1) sensitive to the spatial position of the group as a whole and 2) sensitive to the spatial positions of particular individuals within the group, and whether it is possible to assess which is the most influential.

MATERIALS AND METHODS

Direction matching for sparse data sets

The force matching method proposed by Eriksson et al. (2010) identifies optimal models of interactions in animal groups by adapting the method used to describe interactions among particles (Ercolessi and Adams 1994). In essence, it develops models that describe how other group members (hereafter "associates") influence a focal animal and then compares how well these models match the observed motion of the focal animal. As the method deals with measures of force (Force = Mass × Acceleration), it relies on knowing the acceleration of each individual at each time point in order to calculate the forces exerted on any one animal by its associates.

In sparsely measured movement data, acceleration cannot be estimated accurately, which presents a problem in applying the force matching method to coarse-grained data sets (Eriksson et al. 2010). If we modify our predictions, however, from those dependent on acceleration to those based solely on the direction of travel—a measure easily estimated for sparse movement data—it is possible to apply a version of force matching that allows us to quantify the influence that others have on the direction of travel of a focal individual (Figure 1): We therefore refer to this as direction matching. It is then possible to test models that describe how the specific locations or movements of others influence the movement of a focal animal. This enables us to address questions about whether individuals show signs of moving to the center of the group, for example, or whether they avoid higher ranking individuals.

Specifically, our proposed procedure follows 3 steps: 1) Identify a focal individual and, at each observed point (x_t, y_t), measure its observed direction of travel to the subsequent observation point $\left(\hat{\boldsymbol{v}}_t = \left(x_{t+1} - x_t, y_{t+1} - y_t\right) / \,||\, x_{t+1} - x_t, y_{t+1} - y_t \,||\right)$ and the direction of travel from the preceding observation ($\hat{\boldsymbol{v}}_{t-1}$); 2) use linear interpolation between sequential observations of all associates to estimate their position and their direction relative to the point of observation of the focal animal ($\hat{\boldsymbol{v}}_{i,j}$) (Figure 1). By using this approach, we generate a data set containing the observed direction of travel associated with a given spatial structure of associates. It is then possible to 3) use this spatial structure to search for a model that minimizes the squared difference between the observed and predicted direction of travel: $\varepsilon = ||\, \hat{\boldsymbol{v}}_{t,obs} - \hat{\boldsymbol{v}}_{t,pred} \,||^2$, where $\hat{\boldsymbol{v}}_{t,\mathrm{obs}}$ is the observed direction of travel at time t and $\hat{\boldsymbol{v}}_{t,\mathrm{pred}}$ represents estimated direction of travel based on hypothesized functional relationships between external factors and direction of travel taken by a focal animal.

There are 3 sources of errors associated with this approach that increase as the data become more sparse. First, when the time elapsed between the current and succeeding observations increases ($\Delta t_{\mathrm{next}} = t_1 - t_0$), the estimated direction of travel will be influenced increasingly by other events that occur between the 2 time points, rather than by only the group structure at time t. Second, as the time elapsed between preceding and current observation points of the focal individual increases ($\Delta t_{\mathrm{prev}} = t_0 - t_{-1}$), the previous bearing will have decreasing influence on the subsequent one. The third source of error results from the interpolation of associates' positions and directions of travel, as the error of both estimates will increase as data become more sparse.

Simulated data set

To test the feasibility of the direction matching approach, we used Repast Simphony (North et al. 2013) to simulate a group of agents ($N = 14$) with fixed behavioral rules (i.e., known behavior). The model was based loosely on the findings of Strandburg-Peshkin et al. (2015), who used a high-resolution data set (every second), in which stationary baboons showed a sensitivity to following moving animals, "initiators," if there were many of them and they were in high DA. The authors quantified DA of the initiators as 1 minus the total circular variation of the angles from a focal individual to each

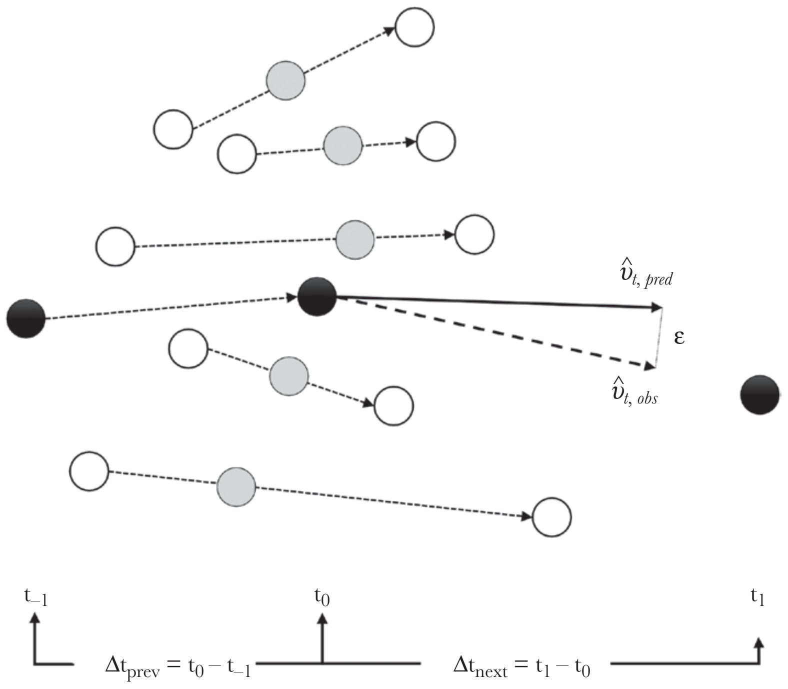

Figure 1

Diagram of data extraction method for coarse data. The focal animal's location is indicated by black circles, whereas gray circles identify the estimated position of other group members at the same observation time. The observed direction of travel ($\hat{\boldsymbol{v}}_{t,\mathrm{obs}}$) between t_0 and t_1 is then contrasted with the modeled direction of travel ($\hat{\boldsymbol{v}}_{t,\mathrm{pred}}$).

initiator (min = 0, max = 1) (Figure 2). To mimic these findings, we developed a simulation model of a group of agents with simple foraging behaviors. These agents actively search/move for food patches located on a 2D landscape and maintain cohesion by moving toward the group once an isolation threshold has been reached (Supplementary Figure S1 and Table S1). This isolation threshold was based on the magnitude of the DA and the sum of inter-individual distances (D) to all associates from the focal agent. DA provided a measure of the relative clustering of associates in terms of direction from the focal agent, and the D threshold captures the relative spread of the group from the focal agent (Figure 2). Agents would thus surpass the isolation threshold when the group was spread out and their directional clustering was high. Specific values for defining this threshold were chosen to produce active movement in the simulated group ($D > \alpha_d = 1000$ m, DA $> \alpha_{da} = 0.85$) (Figure 2). Individuals in the group were also given a preference for one individual in the group, such that, when they were not hungry, the individual moved toward a prespecified associate within the group when that associate was not within a social radius (α_s). Preferences in the simulated group (i.e., who moves toward whom in the group) were assigned in a way that allowed us to test the abilities of the direction matching method to extract attraction coefficients from different attraction structures: 1) chains of attractions, 2) triadic attractions, and 3) reciprocal attraction structures (presented in the results section below). This simulation generated group movement data in which both anti-isolation behaviors and preferential movement behaviors toward specific individuals in the groups were present, providing a test of the abilities of the direction matching approach to identify the correct behaviors taking place within the group. This simulation generated movement data in which behavior was not constant: Foraging individuals moved in a straight line or toward food patches when these were available, and only moved based on the position of other group members when the isolation threshold was exceeded or when not hungry. We simulated a group of 14 individuals for a period of 48 h, recording every agent's position at 1-s intervals.

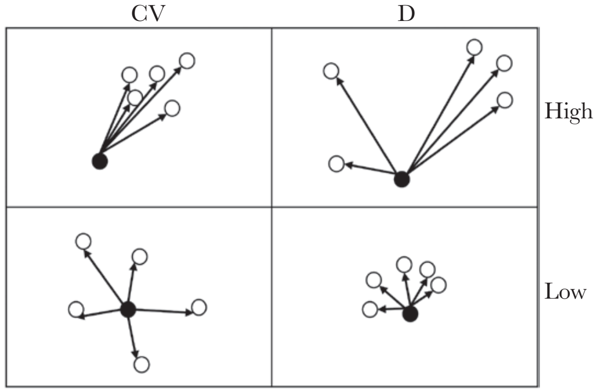

Figure 2

Diagram of DA and total inter-individual distance (D). Both measures are egocentric, focusing on a focal animal.

We generated additional data sets by progressively degrading the simulated data by sampling from a distribution of increasingly sparse revisit times. We used our own observed data on the movement patterns of baboons (see below) as a base estimate of the general structure of revisit times (i.e., the time elapsed between data points recorded for a given animal) captured in the field by a single observer, where the best fit distribution by maximum likelihood was lognormal with shape = 6.1 and scale = 0.6. We altered this distribution by shifting the shape parameter to produce a range of revisit time distributions, varying from more to less frequent revisits (i.e., mean revisit times of 1, 5, 10, 15, and 20 min). Using these probability distributions, we resampled from the full simulated data set (1 s) to generate data sets of increasingly dispersed observations, mimicking variation in capture times (Figure 3), containing respectively: 172 775 (1 s), 3247 (1 min), 610 (5 min), 323 (10 min), 212 (15 min), and 151 (20 min) observations. Additionally, we quantified the effect of increasing the duration over which observations were recorded (referred to as the "extent" of the data below) to assess whether increasing sample size in this way could compensate for reduced data resolution. To do so, we simulated additional movement data sets to provide 48, 96, 144, and 192 h of movement data with a temporal resolution of 10-min mean revisit times, containing respectively 323 (48 h), 662 (96 h), 967 (144 h), and 1310 (192 h) observations.

Baboon data set

Our field data set consisted of 74 days of full-day follows of a baboon troop at the De Hoop Nature Reserve in South Africa (Barrett et al. 2004). Individual GPS points of all adult group members ($N = 14$) were collected continuously throughout the day by an observer walking repeatedly from one end of the group to the other (Andrienko et al. 2013; Dostie et al. 2016). A GPS point was taken on all adults present in the group by holding the GPS receiver above the animal (or as close as possible) to record its position and individual identity. In field tests, the GPS receivers were found to be accurate to within 2–5 m. This generated 61 842 points, with a mean revisit time for each individual of 9 min (median 7 min). The distribution of revisit times was lognormal shaped (shape = 6.1 and scale = 0.6), with a long right hand side tail, where 90% of all revisit times were less than 15 min. For our sample focal animal from this group—AS, the second highest ranking female—this provided 4998 observation points covering 708 h. As the points were not collected concurrently, linear interpolation between observation points was used to estimate associates' positions at the observed times of AS (see *Direction matching for sparse data sets* section).

ANALYSIS

Direction matching models tested

We propose 2 models based on an individual's response to group isolation and social isolation, respectively (Equations 1 and 2). These allow us to test 2 hypotheses concerning the way that animals are thought to maintain cohesion within a group: one in which the animal is largely responding to the group as a set of homogenous individuals and the other where individuals respond to the group as a heterogeneous collection of individuals. Equation 1 assumes that the motion of the focal animal is solely influenced by the group, whereas Equation 2 includes the influence of both the group as a whole and individuals within the group. By incorporating both hypotheses within Equation 2, we can compare between these 2 hypotheses.

$$\hat{\boldsymbol{v}}_t = \beta_b \hat{\boldsymbol{v}}_{t-1} + \left(\beta_{\mathrm{cm}} \hat{\boldsymbol{v}}_{\mathrm{CM}} \mid D_t > \alpha_{\mathrm{d}} \text{ and } \mathrm{DA}_t > \alpha_{\mathrm{d}}\right) \tag{1}$$

$$\hat{\boldsymbol{v}}_t = \beta_b \hat{\boldsymbol{v}}_{t-1} + \left(\beta_{\mathrm{cm}} \hat{\boldsymbol{v}}_{\mathrm{CM}} \mid D_t > \alpha_{\mathrm{d}} \text{ and } \mathrm{DA}_t > \alpha_{\mathrm{da}}\right) + \sum_{i \neq j}^{n} \left(\beta_{i,j} \hat{\boldsymbol{v}}_{i,j} \mid D_{i,j,t} > \alpha_{\mathrm{s}}\right) \tag{2}$$

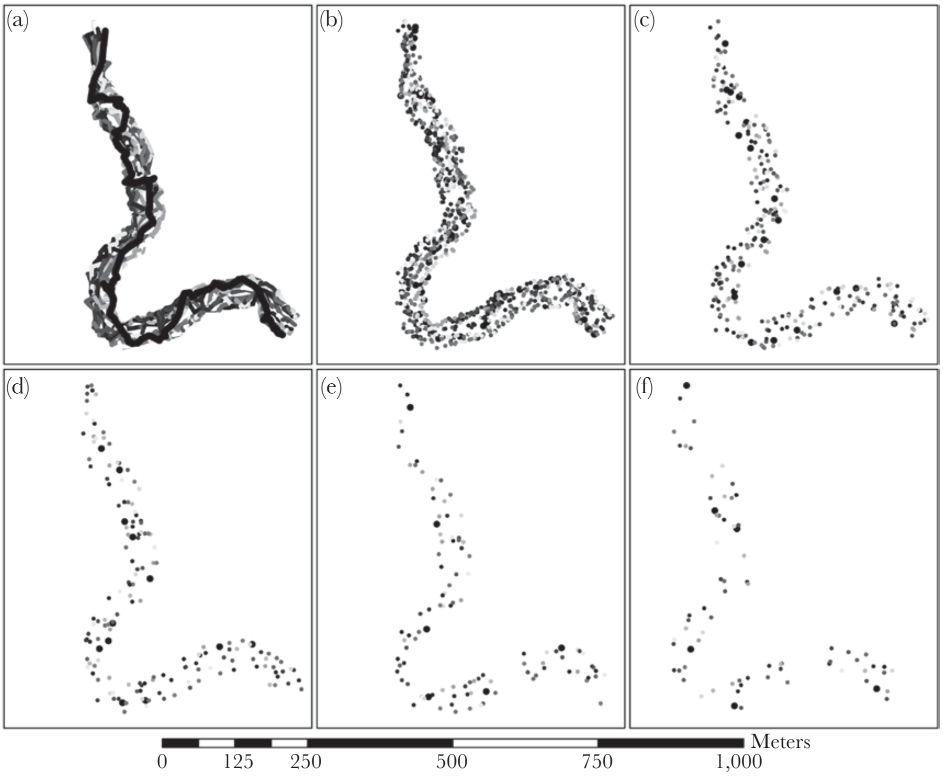

Figure 3

(a) Positions of simulated agents over a 100-min period, with increasingly sparse observations: (b) 1-min, (c) 5-min, (d) 10-min, (e) 15-min, and (f) 20-min mean revisit times. The thick black line/point represents the position of the focal individual on which the direction matching was applied, associates are assigned smaller points with gray shading.

where $\hat{\boldsymbol{v}}_t$ is the observed travel direction of the focal animal i and is influenced by external factors: $\hat{\boldsymbol{v}}_{t-1}$ is the previous direction of travel (i.e., previous bearing), $\hat{\boldsymbol{v}}_{\mathrm{CM}}$ is the circular mean of the directions to all associates from the focal animal i, and $\hat{\boldsymbol{v}}_{i,j}$ is the direction to individual j from focal animal i (Figure 4). The parameters β_{b}, β_{cm}, and $\beta_{i,j}$ represent the weights of each influencing factor, and each provides the strength of its predictive influence on the focal individual's direction of travel. Parameters α_{d}, α_{da}, and α_{s} are estimates of the conditions in which a focal individual is considered isolated, for example, a highly spread group (α_{d}) in which DA is high (α_{da}) and/or the distance from preferred group associates is high α_{s}. We measured group isolation of the focal animal i at time t in terms of DA in associates (DA_t) and the spread of associates (D_t) relative to the focal individual. Social isolation of the focal individual i at time t was measured as the straight line distance between 2 individuals ($D_{i,j,t}$).

Fitting direction models

We used GeoTools in java (www.geotools.org) to develop a tool that extracts the direction of travel of a focal individual, the position and direction of travel of associates, the mean circular angle, and inter-individual distances at each observation time. The output from this tool can then be treated as an overdetermined system of equations, describing changes in x and y positions. In the case of linear equations, it is then possible to fit parameters by a least squares method ($\boldsymbol{X} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{Y}$). Alternatively, with nonlinear equations, such as Equations 1 and 2, optimization methods can be used. We used the DEoptim algorithm (Mullen et al. 2011) in the R environment (R Core Team 2015) to minimize the sum of squares difference between observed and predicted direction of motion. As our equation is linear with conditional statements, we used the DEoptim algorithm to select values for the conditional parameters (α_{d},α_{da}, and α_{s}). We then used nonnegative least squares (Soetaert et al. 2009) to solve the remaining linear equations and return the sum of squares difference between observed and predicted directions of travel to the DEoptim algorithm. Nonnegative least squares were used as we are interested only in attractive influences in our hypothesis. This optimization method reduces the parameter space explored by DEoptim (i.e., parameter space of 3) and takes advantage of the linear aspects of Equations 1 and 2. For the simulated data sets, the range explored by DEoptim of the 3 conditional parameters was set at 0–5000 m for α_{d}, 0–1 m for α_{da}, and 0–100 m for α_{s}. For the baboon data set, we kept the range of conditional parameters the same as in the simulated case, but increased the range for social distance (α_{s}) to 0–1000 m, as we are less certain about the potential range of this parameter. As autocorrelation was present in the residuals of models for higher frequency data sets, all confidence intervals (CIs) for attraction coefficients were estimated using a block bootstrapping approach, using 1 h as the geometric mean value of block length (Politis and Romano 1994; Canty and Ripley 2016).

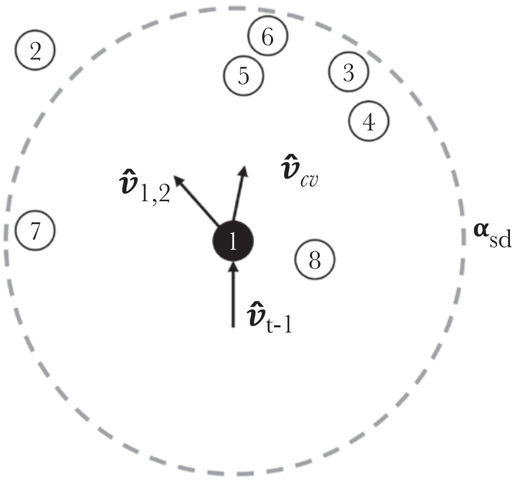

Figure 4
Visual diagram of Equation 2 applied to one focal baboon (AS).

RESULTS

Simulation results

Effect of temporal resolution

When the frequency of the simulated data was at the maximum temporal resolution (i.e., 1-s observations), the direction matching approach successfully identified the influence of the group as a whole on focal animal (agent 0) movements ($\beta_{\mathrm{cm}} = 0.85$) (Figure 5a). Similarly, estimates of the influence of associates on direction of travel were all 0, except for the one individual (agent 2) to which the focal animal was attracted in the simulation ($\beta_{0,2} = 0.001$). However, the 95% CI for $\beta_{0,2}$ contained 0 (95% CI: 0, 0.004) suggesting uncertainty in this estimate. Not surprisingly, the estimated coefficient of the previous direction of travel was also found to have a strong influence on the observed direction of travel ($\beta_{\mathrm{b}} = 0.91$). Furthermore, the method was able to identify the correct conditional parameters defining when anti-isolation behaviors, that is, movement toward the group mean direction or movement toward a social partner, would occur: $\alpha_{\mathrm{iid}} = 350.12$, CI = 349.83, 350.39; $\alpha_{\mathrm{da}} = 0.80$, CI =0.801, 0.800; $\alpha_{\mathrm{s}} = 3.9$, CI = 0.0, 8.2, with the social parameter estimate α_{s} containing 0 within the estimate of the 95% CI. As we resampled the data to lower temporal resolutions, we quickly lost accuracy in the estimates of α_{d}, α_{da}, and α_{s} (Figure 6). More promisingly, however, the ability to distinguish between competing hypotheses was much less sensitive to reductions in data frequency (Figure 5): Specifically, the estimated coefficients for direction to group as a whole ($\hat{\boldsymbol{v}}_{\mathrm{CM}}$) were consistently present, with larger uncertainty around the estimate as temporal resolution was lowered. Estimated coefficients for direction to associates quickly lost the ability to distinguish between influence from the true attraction to agent 2 and influence from agent 3 (Figure 5). In the simulation, agent 2 was attracted to agent 3, forming a link: 0 → 2 → 3. The estimated coefficients consistently distinguished agents 2 and 3 from influence of all other agents up to 10- to 20-min temporal resolutions, where the estimated coefficients suggest attraction to agent 1 (in the simulation, agent 1 is attracted to the same agent as agent 0: i.e., 1 → 2).

Effect of temporal extent

When the time frame over which the data were analyzed was increased from 48 to 96, 144, and 192 h, we found that the estimates of coefficients and conditional parameters became more accurate (Figures 7 and 8). Estimates of conditional parameters gained only marginal improvements, whereas at 192 h of observation, estimates of the coefficients for direction toward specific associates ($\hat{\boldsymbol{v}}_{i,j}$) more confidently identify agents 2 and 3 as having an influence on motion of the focal agent 0.

Estimating attractions within the group

When the estimated attractions between individuals in the group were compared using the 1-s data set and the 10-min data set

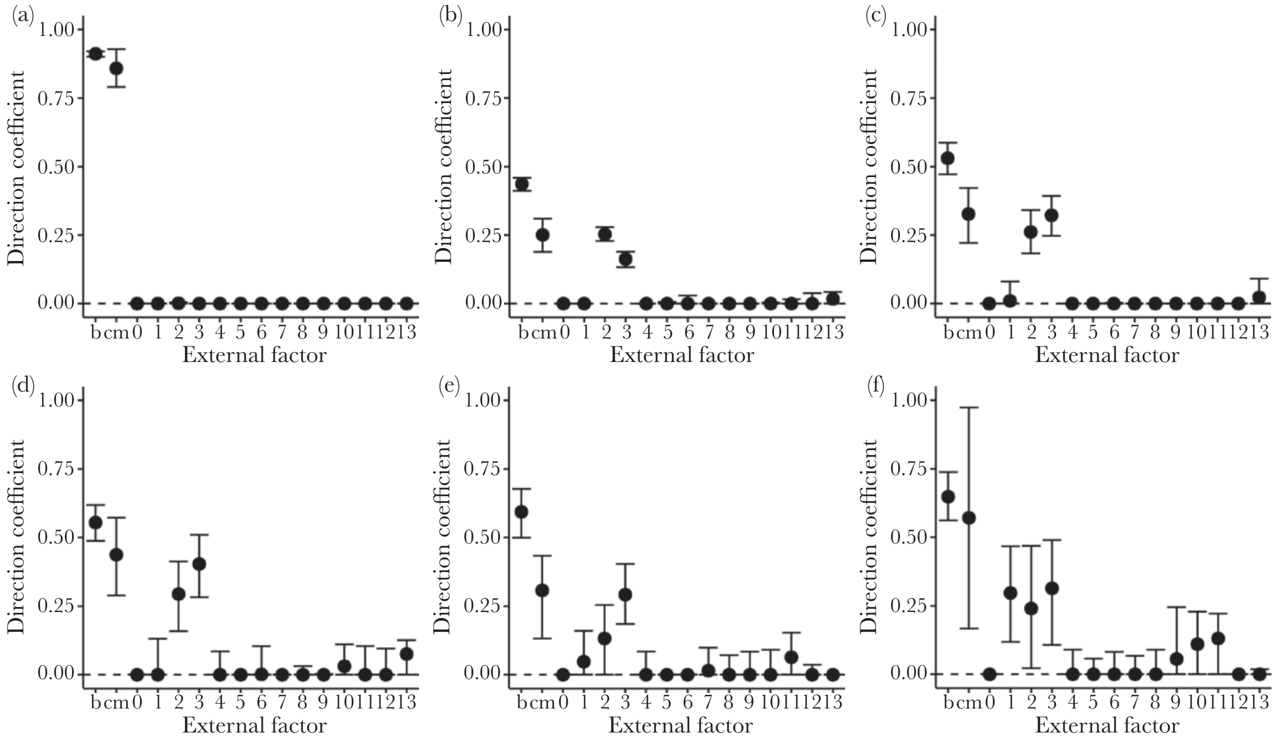

Figure 5

Estimates of influence for the (a) nonreduced data set, (b) 1-min, (c) 5-min, (d) 10-min, (e) 15-min, and (f) 20-min mean revisit times. External factor "b" represents the influence of the previous bearing, "cm" is the influence toward the circular mean, and the numbered factors represent the influence toward each group mate.

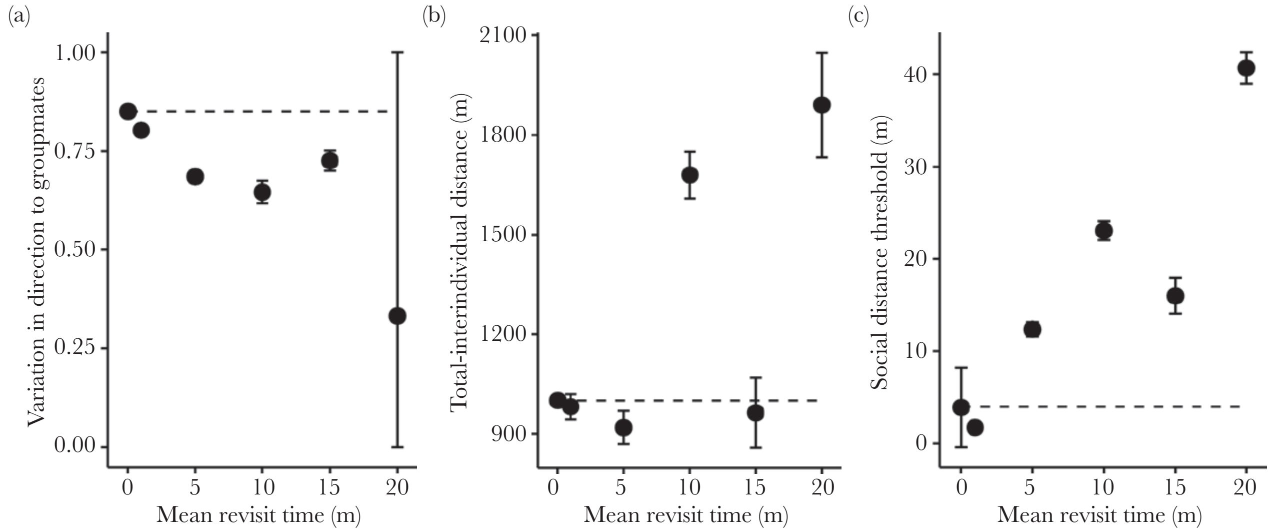

Figure 6

The estimated conditional parameters as the temporal frequency of movement data decreases: (a) threshold of isolation due to the magnitude of DA in associates (α_{da}), (b) threshold of isolation due to inter-individual distances from the focal animal (α_d), and (c) threshold of isolation from social partner (α_s). Error bars represent 95% CI for each parameter estimate, dashed lines are the true values.

against the true (simulated) attractions, we found that estimates produced a different pattern of attractions depending on the resolution of the movement data (Figure 9). The low frequency data set (10 min) produced estimates that captured the general flow of attraction, with individuals showing estimates of moving toward individuals "downstream" of their true social attractions. For example, in the simulation, agent 4 moved toward agent 5 when not hungry. When using sparse data, the direction matching method estimated positive influence coefficients to 6 and 12 as well as 5. This is, however, a sense in which all these individuals form a "chain" of attraction, that is, in the simulation agent 4 is attracted to 5, 5 to 6, and 6 to 12.

The high-frequency data (1 s), in contrast, showed the ability to correctly identify the specific nature of these individual attractions.

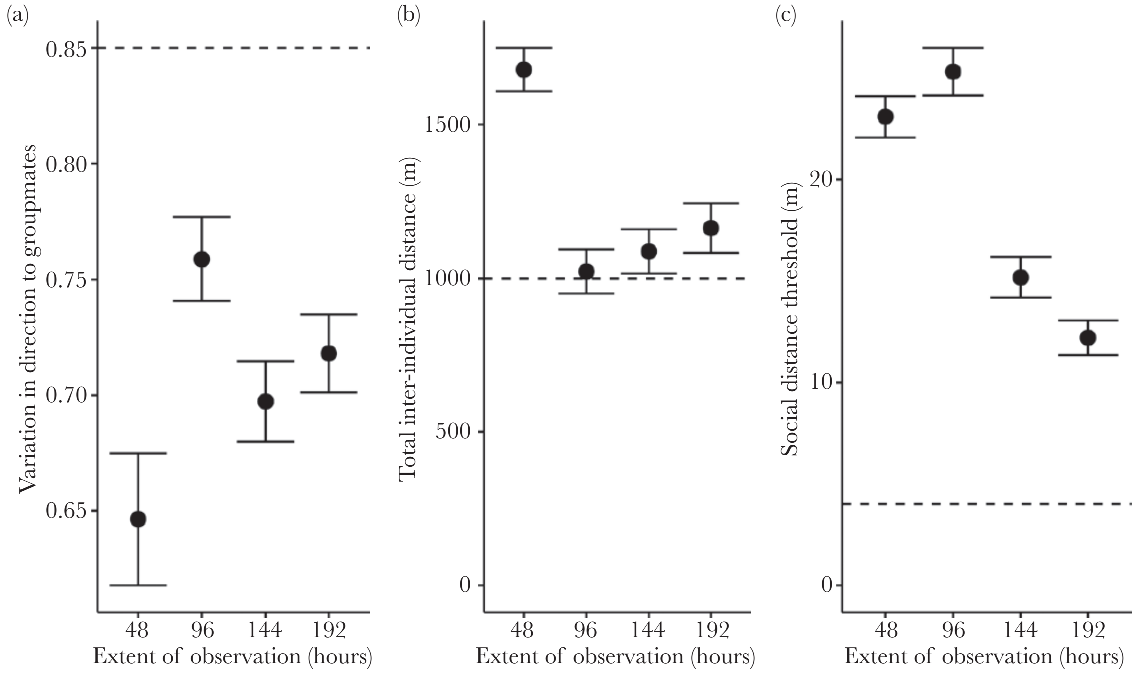

Figure 7

Conditional parameter estimates are presented as the temporal extent of the observed data was increased from 48 to 192 h: (a) threshold of isolation due to the magnitude of DA in associates (α_{da}), (b) threshold of isolation due to total inter-individual distance from the focal animal (α_d), and (c) threshold of isolation from social partner (α_s). The dashed horizontal lines represent the true values, and the vertical bars are the 95% CIs.

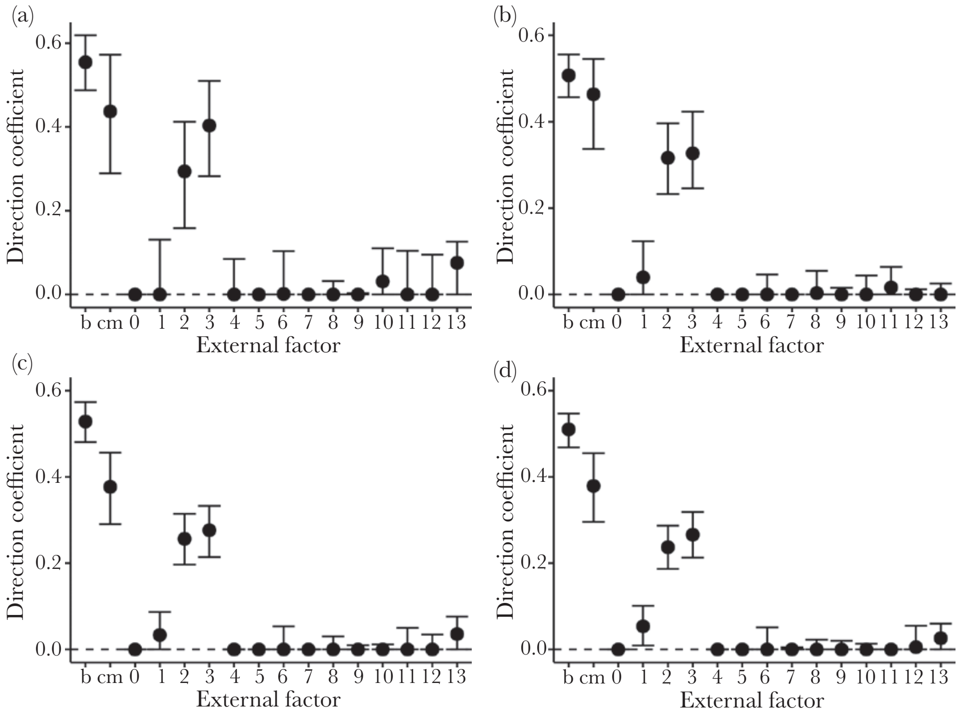

Figure 8

The estimated influence coefficients as the extent of data is increased from (a) 48 h, (b) 96 h, (c) 144 h, and (d) 192 h. Estimated coefficient for previous direction of travel (b), direction toward the mean group direction (cm), and direction from the focal agent 0 to each agent in the group are presented along the *x* axis. Vertical bars are the 95% CIs.

Nevertheless, there was a clear effect whereby individuals that performed fewer social movements in the simulation, that is, those at the center or downstream of the flow of attraction, and so less socially isolated due to the attractions of others, were found to have smaller estimated influence coefficients ($r = 0.9$, $P < 0.001$). This may be a consequence of the fact that, as the recording frequency

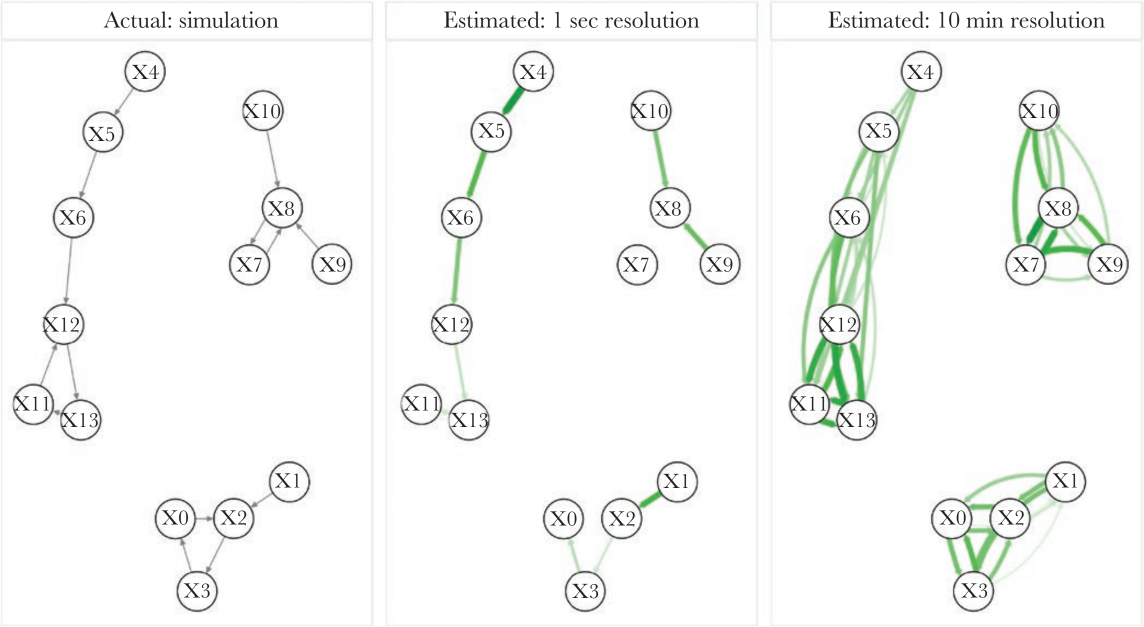

Figure 9
Attraction networks for the true attractions implemented in the simulation, estimated influence coefficients using high-frequency data (1-s revisit time), and estimated influence coefficients using sparse movement data (10-min revisit time). Revisit time refers to the average time between observations of an individual in the group, in which location and id were recorded.

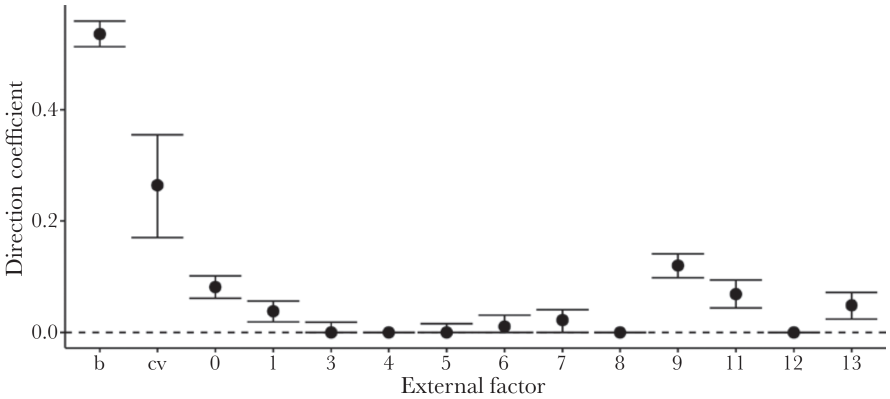

Figure 10
The estimated influence from one focal baboon in the observed GPS data. External factor "b" represents the previous bearing, "cm" represents the pull toward the mean circular direction of the group from the focal individual, and the numbered factors represent the pull toward specific individuals in the group.

of the data increased, the relative proportion of observed directions of travel that were influenced by movement toward a social partner was reduced. Over a period of 1 h, for example, observed directions of travel at intervals of 1 s might be the result of attraction toward a social partner only a few times, whereas a 5-min revisit time will generate a relatively higher proportion of observed directions of travel due to attractions toward a social partner.

This can be seen in the increase in the magnitude of the uncertainty around the estimated coefficients as temporal resolution is decreased: At 1 s, the difference between the estimate and its associated 95% upper CI (standardized by the estimate: i.e., ($\beta_{0,2}$ − upper CI) / $\beta_{0,2}$) is 2.6, at 1 min 0.1, at 5 min 0.3, at 10 min 0.4, at 15 min 0.9, and at 20 min 1.0, that is, we get a U-shaped curve. As the 95% CI was calculated via bootstrapping, we would expect that at 1-s intervals, where many of the observed directions of travel are not based on direction to social partners, random sampling would increase the uncertainty of the estimate. At decreased temporal resolutions, more of the observed directions of travel are related to movement to social associates, thus reducing the uncertainty of the estimate. As the temporal resolution is further decreased, we again see an increase in uncertainty due to the fact that, with longer periods between observations, there are increasing numbers and sequences of behaviors. In the case of our simulated population, 1 min would seem the optimal point at which to collect data in order to reduce the uncertainty around these social attraction estimates. These effects are especially clear in the cluster of agents

11, 12, and 13 (i.e., a triadic attraction structure), where agents upstream of the cluster show increasing estimates of influence compared with those at the center (Figure 9). In this cluster (11, 12, and 13), the estimated influence coefficients are an order of magnitude lower, with the influence coefficient between 11 and 12 being correctly identified as greater than 0, but containing 0 in its 95% CI. Similarly, this is the case with the reciprocal attractions between individuals 7 and 8, as well as the other triadic structure containing individuals 0 and 2 (Figure 9).

Baboon results

Applying the direction matching method to a sample focal animal in our real baboon data set, using Equation 1, we found that AS was sensitive to the previous bearing ($\hat{\boldsymbol{v}}_{t-1}$) and the position of the group as a whole ($\hat{\boldsymbol{v}}_{\mathrm{cm}}$). More specifically, the model suggested that, when DA in the group was above 0.13 and the spread of the group was larger than 5.1 m, the focal animal was influenced by mean group direction. These conditions apply in 96% of the observations, indicating a wide-spread influence of the group on travel direction.

When we applied Equation 2 to the baboon data set in order to test between the 2 hypotheses of cohesion (i.e., sensitivity to the group vs. sensitivity to specific individuals), we found that AS's direction of travel showed signs of being sensitive to previous bearing ($\hat{\boldsymbol{v}}_{t-1}$), the group as a whole ($\hat{\boldsymbol{v}}_{\mathrm{cm}}$), and specific individuals within the group ($\hat{\boldsymbol{v}}_{i,j}$) (Figure 10). The largest influence on direction of travel was that of the previous heading of the animal ($\hat{\boldsymbol{v}}_{t-1}$). When we examine the estimated isolation conditions under which the $\hat{\boldsymbol{v}}_{\mathrm{cm}}$ and $\hat{\boldsymbol{v}}_{i,j}$ influenced travel direction, we found that $\hat{\boldsymbol{v}}_{\mathrm{cm}}$ was estimated to become a factor only when the magnitude of DA was greater than 0.94 and where the sum of the inter-individual distances (D) was not influential (95% CI = 0, 15.6 m). The DA condition (i.e., DA > 0.94) occurred in only 4.2% of the data, suggesting that moving toward the group as a whole was infrequent and occurred only when DA was very high. In terms of moving toward particular animals in the group, AS was sensitive (in order of highest to lowest effect) to the highest ranking female (SA, id 9), the alpha male (SC, id 11), the third highest ranking female (VI, id 13), and a low ranking female (AC, id 1) (Figure 10). She also showed signs of attraction toward a transient subadult male (KN, id 0). AS was estimated to be attracted to these specific individuals when their distance was more than 5.6 m (95% CI = 1, 10.2 m). All estimates of attraction coefficients to other animals in the group had 95% CIs that included 0. Calculating the proportion of variation in direction of travel explained by each model, as 1 minus the residual sum of squares over the total sum of squares, we find that the r^2 of Equation 1 was 0.30, whereas the r^2 of Equation 2 was 0.32.

DISCUSSION

Our results indicate that sparse data are able to estimate functional relationships between specific contexts and particular movement behaviors and thus can be used to interrogate the individual movement behaviors that underpin the coordinated movement of social groups. Although we lost accuracy, as well as the power, to identify the correct model structure as the data became increasingly sparse, this was ameliorated by increasing the extent of the data set. At present, extended periods of observations using animal-mounted products are limited by battery life, whereas data collection by field researchers, at least in principle, offers the possibility of long-term observations, such that collection of sparse movement data may be more feasible for many field researchers. In addition, given that increasing the extent of the observational period will almost certainly result in capturing a broader range of social and ecological conditions, models fitted to increasingly large data sets should also provide more generalized results. This suggests that the direction matching approach with sparse data is well suited to detecting common and consistent patterns of motion. This is useful for questions concerning behavioral responses to frequently encountered social and ecological conditions (e.g., dyadic interaction within groups, common group formations/structures, or reducing isolation), but less useful for identifying specific behavioral responses to rare events, where higher frequency data and a focus on shorter time scales are likely necessary. Our simulation also suggested that parameter estimation of the contexts under which a behavioral pattern occurred will similarly benefit from high-resolution data. Importantly, however, comparisons that assessed the relative magnitude of influential factors were less sensitive to data resolution (Figure 5).

Estimating attractions between individuals in the group using direction matching suggests that there is a tradeoff when varying the temporal resolution of the observed data set. When the resolution of the data set was reduced (average 10 min between points), the direction matching method was able to estimate the general flow of attraction between individuals in the group, whereas the high-frequency data set direction matching was able to correctly identify specific between-individual attractions, albeit at a cost of increased sensitivity to the proportion of time individuals spent performing this attraction behavior. This sensitivity resulted from increased variation in movement behavior as temporal resolution was increased.

One way to think about these results, then, is to envisage adjusting the focus on a microscope: At sparse temporal resolutions, we can extract a group-level view of inter-individual influence patterns, but as we adjust the focus and increase temporal resolution, we achieve a more detailed account of the influence between specific individuals in the group, at the cost of increased variation or noise. That is, estimates of influence using coarse movement data sets are more heavily influenced by the attraction patterns of others in the group (e.g., 4 attracted to 5, 5 attracted to 6, etc.) which therefore provides a group-level view of social attractions whereas, for data sets with increased temporal resolution, these effects are reduced, which allows the identification of individual-level estimates (Figure 9).

As individuals cluster together spatially in the simulation due to their specific attractions, it is perhaps not surprising that the ability to estimate attractions between specific individuals within clusters is reduced when using sparse data. This suggests that the degree to which a population is spatially clustered will be an important factor in the ability of the method to identify the correct influence when using increasingly sparse data sets. For example, a group that is consistently tightly clustered will have significantly less variation in the direction of groupmates, limiting the ability of direction matching to estimate attraction weights. In such a case, we would expect that higher frequency follows would be required to achieve specific between-individual estimates. Conversely, in groups that are more widely dispersed, with a high degree of within-group structure, the use of direction matching is likely to be robust to reductions in temporal resolution of the movement data set. This suggests that the overall spatial geometry of a social group is an important consideration when choosing a resolution at which to collect data on individual movement patterns and applying direction matching to gain insight into inter-individual influence patterns.

When applied to a focal individual, the direction matching approach found consistent patterns that suggested preferential movement toward particular individuals. Specifically, our focal animal was most sensitive to location of the highest ranking females (SA = rank 1 and VK = rank 3) and the alpha male (TR), which accords with social association data collected independently of the GPS values (Barrett L, Henzi SP, unpublished data). Given the sparse nature of our data set, we treat these estimates of moving toward specific individuals as the consequence of other groupmate attractions as well, thereby proving an estimate of the flow of attraction in the group. As found elsewhere, with high-resolution data (Strandburg-Peshkin et al. 2015), the mean group direction was also influential in explaining the observed motion of a single animal, although this only emerged when the DA of the group was high (i.e., highly clustered in one direction from the focal animal).

By assessing multiple possibilities within a single model, the direction matching approach provides a tool that can help tease apart the relative influence of a range of potential factors in explaining observed motion. Complicating the comparisons between the tendencies to move in the mean group direction versus attraction to specific individuals, however, is their relatedness. For example, the correlation between mean direction of the group and the direction to the highest ranking female (SA) relative to the focal animal (AS) was $r = 0.59$ ($t = 71.66$, $df = 9610$, $P < 0.001$). Nevertheless, when both mean group direction and individuals considered (Equation 2), we found that AS was most influenced by SA, although SA was not the animal whose movement was most closely aligned to the mean group direction (this was VK, with a value of $r = 0.69$). AS's movement therefore appeared to be more sensitive to that of SA, rather than to the mean direction of the group, so it does seem possible to tease apart these different patterns despite the correlation between the 2 rules. The results from Equation 2 suggest, though, that when the DA of the group was high, the mean direction of the group became more influential than the direction of any particular individual (Figure 10), although it is clear that such conditions occurred only infrequently (4.2% of the data).

Given that force matching is derived from physics and deals with particle interactions, rather than the behavior of animate agents, it raises an interesting issue regarding interaction among animals, and what is meant by "force." Along with changing the name to direction mapping, we also suggest relaxing the assumption that individual animals will always react instantaneously to all changes in their associates' positions (as might be expected of a physical system). We assume, instead, that certain factors, such as those we consider here (associates' positions and mean direction of the group), will introduce a consistent bias in observed motion. Direction matching can then be used to estimate this bias and the conditions under which it emerges (e.g., the mean direction of the group only becomes influential when DA in the group is high, as we found there), and such estimates can then be used to test between competing hypotheses about the factors that are driving observed motion.

Summary

Our results suggest that, within determinable limits, our proposed modification to the method of force matching (i.e., direction matching) can be used successfully to extract patterns of individual movement from temporally sparse movement data. We demonstrate this using simulation models to recover a prespecified pattern of behavior. Applying our approach to data from wild baboons, we were able to identify consistent behavioral responses in a focal baboon in relation to the movements of others. Given the wide availability and low cost of handheld GPS devices, we suggest that sparse movement data sets of social groups can provide a valuable means for developing and empirically testing models of how individuals control their local social and ecological environments.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://www.beheco.oxfordjournals.org/

FUNDING

This work was supported by the Leakey Foundation (San Francisco, CA); National Research Foundation (Pretoria, South Africa); and Natural Science and Engineering Research Council of Canada (NSERC) grants to S.P.H. and L.B. L.B. is also supported by NSERC's Canada Research Chairs Program (Tier 1). T.R.B. is supported by a Fonds de Recherche Nature et Technologies (FQRNT) Postdoctoral Fellowship and funds from L.B.'s Canada Research Chair award.

We thank Dr Parry Clark who collected the baboon spatial data, Dr Julie Teichroeb and Dr Rafael Reyna-Hurtado who provided helpful comments on the manuscript, 3 anonymous reviewers who provided insightful suggestions, and Mel Lefebvre who provided editorial comments. The running of optimization algorithms was facilitated through the use of Westgrid/Compute Canada facilities. Cape Nature gave permission for the baboon research.

Data accessibility: Analyses reported in this article can be reproduced using the data provided by Bonnell et al. (2016).

Handling editor: Shinichi Nakagawa

REFERENCES

Andrienko N, Andrienko G, Barrett L, Dostie M, Henzi P. 2013. Space transformation for understanding group movement. IEEE Trans Vis Comput Graph. 19:2169–2178.

Aureli F, Schaffner CM, Asensio N, Lusseau D. 2012. What is a subgroup? How socioecological factors influence interindividual distance. Behav Ecol. 23:1308–1315.

Barrett L, Gaynor D, Rendall D, Mitchell D, Henzi S. 2004. Habitual cave use and thermoregulation in chacma baboons (*Papio hamadryas ursinus*). J Hum Evol. 46:215–222.

Bonnell TR, Barrett L, Henzi P. 2016. Data from: direction matching for sparse movement datasets: determining interaction rules in social groups. Dryad Digital Repository. http://doi: 10.5061/dryad.kv2kh.

Cagnacci F, Boitani L, Powell RA, Boyce MS. 2010. Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. Philos Trans R Soc B Biol Sci. 365:2157–2162.

Canty A, Ripley BD. 2016. boot: Bootstrap R (S-Plus) functions. R package version 1.3-18.

Conradt L, Roper TJ. 2005. Consensus decision making in animals. Trends Ecol Evol. 20:449–456.

Dalziel BD, Le Corre M, Côté SD, Ellner SP. 2015. Detecting collective behavior in animal relocation data, with application to migrating caribou. Methods Ecol Evol. 7:30–41. doi:10.1111/2041-210X.12437

De Groeve J, Van de Weghe N, Ranc N, Neutens T, Ometto L, Rota-Stabelli O, Cagnacci F. 2015. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods (SAM). Methods Ecol Evol. 7:369–379.

Ercolessi F, Adams JB. 1994. Interatomic potentials from first-principles calculations: the force-matching method. Europhys Lett. 26:583.

Eriksson A, Nilsson Jacobi M, Nyström J, Tunstrøm K. 2010. Determining interaction rules in animal swarms. Behav Ecol. 21:1106–1111.

Gurarie E, Andrews RD, Laidre KL. 2009. A novel method for identifying behavioural changes in animal movement data. Ecol Lett. 12:395–408.

Handcock RN, Swain DL, Bishop-Hurley GJ, Patison KP, Wark T, Valencia P, Corke P, O'Neill CJ. 2009. Monitoring animal behaviour and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing. Sensors. 9:3586–3603.

Hebblewhite M, Haydon DT. 2010. Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. Philos Trans R Soc B Biol Sci. 365:2303–2312.

Heesen M, Macdonald S, Ostner J, Schülke O. 2015. Ecological and social determinants of group cohesiveness and within-group spatial position in wild assamese macaques. Ethology. 121:270–283.

Kays R, Crofoot MC, Jetz W, Wikelski M. 2015. Terrestrial animal tracking as an eye on life and planet. Science. 348.

King AJ, Douglas CMS, Huchard E, Isaac NJB, Cowlishaw G. 2008. Dominance and affiliation mediate despotism in a social primate. Curr Biol. 18:1833–1838.

King AJ, Sueur C, Huchard E, Cowlishaw G. 2011. A rule-of-thumb based on social affiliation explains collective movements in desert baboons. Anim Behav. 82:1337–1345.

Krause J, Krause S, Arlinghaus R, Psorakis I, Roberts S, Rutz C. 2013. Reality mining of animal social systems. Trends Ecol Evol. 28:541–551.

Laube P, Imfeld S, Weibel R. 2005. Discovering relative motion patterns in groups of moving point objects. Int J Geogr Inf Sci. 19:639–668.

Lukeman R, Li Y-X, Edelstein-Keshet L. 2010. Inferring individual rules from collective behavior. Proc Natl Acad Sci USA. 107:12576–12580.

Dostie MJ, Lusseau D, Bonnell T, Clarke P, Chaplin G, Kienzle S, Barrett L, Henzi P. 2016. Proof of principle: the adaptive geometry of social foragers. Anim Behav. 119:173–178.

Mullen K, Ardia D, Gil DL, Windover D, Cline J. 2011. DEoptim: an R package for global optimization by differential evolution. J Stat Softw. 40:1–26.

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE. 2008. A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci USA. 105:19052–19059.

North MJ, Collier NT, Ozik J, Tatara ER, Macal CM, Bragen M, Sydelko P. 2013. Complex adaptive systems modeling with repast simphony. Complex Adaptive Syst Model. 1:1–26.

Politis DN, Romano JP. 1994. The stationary bootstrap. J Am Stat Assoc. 89:1303–1313.

R Core Team. 2015. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Soetaert K, Van den Meersche K, van Oevelen D. 2009. limSolve: solving linear inverse models. R package version 15. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://CRANR-projectorg/package=limSolve.

Strandburg-Peshkin A, Farine DR, Couzin ID, Crofoot MC. 2015. Shared decision-making drives collective movement in wild baboons. Science. 348:1358–1361.

Sugiura H, Shimooka Y, Tsuji Y. 2011. Variation in spatial cohesiveness in a group of Japanese Macaques (*Macaca fuscata*). Int J Primatol. 32:1348–1366.