A Statistical Framework for Inferring the Influence of Conspecifics on Movement Behaviour

María del Mar Delgado^{1,2*}, Vincenzo Penteriani², Juan Manuel Morales³, Eliezer Gurarie⁴ and Otso Ovaskainen¹

¹*Department of Biosciences, University of Helsinki, Viikinkaari 1, 00014 Helsinki, Finland*

²*Department of Conservation Biology, Estación Biológica de Doñana, CSIC, c/Américo Vespucio s/n, 41092 Seville, Spain*

³*Ecotono, INIBIOMA—CONICET, Universidad Nacional del Comahue, Quintral 1250, 8400 Bariloche, Argentina*

⁴*Department of Statistics, Center for Quantitative Science, University of Washington, Seattle, Washington 98195, USA*

*Correspondence author. E-mail: maria.delgado@helsinki.fi

Summary

1. The movements of individuals - at almost any scale - are likely to depend on the behaviour of conspecifics. As an example, the movements of dispersing juveniles and their settling decisions may depend on the availability of mates and free territories, i.e. both the presence and absence of other individuals. As another example, individuals can use the presence of conspecifics during foraging movements as an indicator of habitat quality.
2. We develop a general statistical framework for identifying and characterizing conspecific influence on movements from tracking data acquired simultaneously from a set of potentially interacting individuals.
3. We model conspecific attraction/repulsion through a functional response in which social behaviour is assumed to depend on proximity to other individuals. The model partitions variation in the functional response into a population component (common to all individuals), variation among individuals (modelled as random intercept-slope), and variation within an individual's trajectory (modelled through temporal autocorrelation).
4. We present a Bayesian approach for the estimation of the model, and illustrate its use with simulated movement data generated from a number of contrasting scenarios. We then apply the method to a case study on eagle owl *Bubo bubo* juvenile dispersal, demonstrating that individual movements are generally influenced by the presence of conspecifics, with the level of attraction decreasing with increasing proximity to other individuals. We further show that female eagle owls are more attracted to conspecifics than males, and both males and females are more attracted to females than to males.

Key-words: behavioural rules, condition-dependent dispersal, mixed-effect model, model/data, social information, statistical inference

1   **Introduction**

2   The presence of conspecifics can influence individual movement patterns, with important consequences
3   in many ecological and evolutionary contexts, such as foraging (e.g. Kawaguchi, Ohashi & Toquenaga
4   2006), mate choice (e.g. Kuussaari *et al.* 1998), habitat selection and space use (e.g. Serrano *et al.*
5   2003). Depending on the relative costs and benefits of conspecific aggregation, individuals may be
6   attracted to or repulsed from other individuals. The strongest benefit of conspecific attraction is
7   expected to take place at intermediate densities, as at low population densities social cues are typically
8   unavailable, whereas at high population densities the potential costs of competition may outweigh the
9   benefits from conspecific attraction (Fletcher 2006; Fletcher 2007). The presence of conspecifics can
10   lead to different behavioural actions, the decision rules behind which can be at least partly inferred
11   from the realized movement patterns (Bode *et al.* 2012). In some cases individuals may primarily
12   follow their nearest neighbours (Hemelrijk *et al.* 2008), or they may use local density as a cue (Delgado
13   *et al.* 2010). In many cases, movement strategies are context-dependent and change over time, either
14   gradually or through discrete behavioural switches (Morales *et al.* 2004).

15   The influence of conspecifics on animal movements has been extensively studied with the help of
16   mathematical models and simulations, especially to understand the underlying mechanisms behind
17   many forms of animal aggregation (e.g. Couzin *et al.* 2002). For example, simple behavioural rules
18   have been shown to lead to patterns of collective motion which are at least qualitatively similar to those
19   observed for flocking birds, schooling fish, or swarming insects (e.g. Couzin *et al.* 2002). Some of
20   these models have been successfully matched to experimental data (e.g. Couzin *et al.* 2011; Herbert-
21   Read *et al.* 2011; Katz *et al.* 2011), but typically not at the temporal resolution at which individuals
22   make their movement decisions. If movement decisions are linked to various types of potentially
23   conflicting motivations (e.g., for food search, predator/competitor avoidance, sampling of potential

nesting habitat, or mate search), they may also be made at different temporal resolutions. In addition, even though movement data are intrinsically complex, with many levels of structure and heterogeneity, previous models have often assumed fairly simplistic rules (e.g. they usually assumed that all individuals operate according to some behavioral rules; Couzin *et al.* 2002; Eriksson *et al.* 2010; Katz *et al.* 2011), with extensions such as context-dependent interactions having only recently emerged (Guttal & Couzin 2010).

In this paper we develop a simple but general statistical framework for studying whether and how the presence of conspecifics influences animal movement at the individual level. We proceed in two steps, first looking for signals of conspecific influence by comparing the observed movement behaviour to that predicted by a null model. We then use a regression model to partition the variation in the patterns of social information use into components corresponding to density-dependent attraction or repulsion, variation among individuals, and variation within individuals. Beside its simplicity, this framework allows incorporating the intrinsic complexity of movement data, and the many levels of structure and heterogeneity in movement behaviour (e.g. across space and time, among individuals, within individuals). We examine the performance of the statistical framework with simulated data and apply the framework to an analysis of eagle owl natal dispersal. The conceptual idea behind the statistical framework is illustrated in Fig. 1.

Materials and methods

NULL MODELS

Whether individuals are attracted to or repulsed from conspecifics can be measured by comparing observed displacements to those predicted by a null model which describes the expectation in case the

1 movements of the focal individual were not influenced by conspecifics (Fig. 1). Let $\mathbf{z}(t) =$
 2 $(x(t), y(t))$ denote the location of the focal individual at time t . Assuming that two consecutive data
 3 points on the location of the focal individual have been acquired at times t_1 and t_2 , we ask if the
 4 observed location $\mathbf{z}(t_2)$ is in some way different from what would be expected from the knowledge of
 5 the location $\mathbf{z}(t_1)$ and the null model of movement between the times t_1 and t_2 . In the ideal case, the
 6 null model should account for all factors influencing movement behaviour except conspecifics, so that
 7 any deviation between the data and the null model could be attributed solely to social factors. While
 8 sophisticated modelling approaches for individual movement have been developed (e.g. Patterson *et al.*
 9 2008), the definition of an appropriate null model is a highly nontrivial task to which we return in the
 10 Discussion. In the analysis of data generated by simulation models (see below), we simply assume that
 11 in the absence of conspecifics, the individual would have taken a movement step with length equal to
 12 that of the observed movement step $|\mathbf{z}(t_2) - \mathbf{z}(t_1)|$, but in a totally random direction (uniform in the
 13 range from 0 to 2π ; a model often referred to as the *simple random walk*). With this choice, the
 14 probability distribution of null locations consists simply of those locations that form a circle around the
 15 previous location $\mathbf{z}(t_1)$, the observed location $\mathbf{z}(t_2)$ being one point on the circle (Fig. 1b).

16

17 MEASURES OF PROXIMITY AND SOCIABILITY

18 After deciding on the null model, one next needs to specify how to compare the observed location
 19 $\mathbf{z}(t_2)$ with the distribution of null locations. To do so, we defined measures of proximity P and
 20 sociability S . We measure proximity P using two alternatives, both of which increase with increasing
 21 local density of individuals. Proximity measure 1 (hereafter, *nearest neighbour*) is defined as the
 22 negative of the shortest distance to any of the other conspecifics. Proximity measure 2 (hereafter,
 23 *density*) is defined as the density (number of individuals per unit area) of conspecifics within a

1 specified radius d from the focal individual. Note that the selection of these indexes (e.g. whether to
2 use the negative of the shortest distance, the inverse of the shortest distance, either of the above for the
3 log-transformed distance, the number of other individuals within a fixed radius, the density of
4 individuals smoothed by a kernel with continuous distance decay, etc.) will in most cases be
5 necessarily somewhat arbitrary due to lack of information to which aspect of proximity the individuals
6 respond to. The selection should be based on biological considerations such as whether the absolute or
7 relative distances are likely to be relevant (whether to log-transform or not), and on statistical
8 considerations, that is, what kind of distributions of the explanatory and / or response variables arise
9 when a given kind of index is applied to the data in hand.

10 We define the sociability measure S as the difference between the observed and null proximity
11 measures: $S = P^{obs} - E[P^{null}]$. Here P^{obs} is the observed proximity of the focal individual to the other
12 individuals after it has taken a movement step, and $E[P^{null}]$ is the expected proximity over the
13 distribution of null locations. Assuming that the individuals make their movement decisions based on
14 the locations of the other individuals at time t_1 the proximity of the location $\mathbf{z}(t_2)$ should be measured
15 against the locations of conspecifics at time t_1 . We denote the sociability measure corresponding to this
16 choice by S^{t_1} . If the individual continuously tracks the other individuals (as may often be the case for
17 real individual movements), the sociability measure should be evaluated by the locations of
18 conspecifics at time t_2 ; we denote this sociability measure as S^{t_2} . The inferential difference between
19 S^{t_1} and S^{t_2} depends on how much the other individuals move during one time step, and hence on the
20 product between the time interval and the movement speed. The measures S^{t_1} and S^{t_2} quantify whether
21 the focal individual at $\mathbf{z}(t_2)$ is closer or further away than expected by the null model from where its
22 conspecifics were at times t_1 and t_2 respectively. A value of $S > 0$ can be interpreted so that the focal

individual was attracted by conspecifics, whereas $S < 0$ indicates that the focal individual avoided conspecifics.

STATISTICAL MODELS

The response variable in our model is S_{it} , i.e. the observed sociability of individual i based on the movement step that started at time t . The main covariate is P_{it} , the proximity of the individual to conspecifics at time t . We normalized P_{it} to zero mean and unit variance, and S_{it} to unit variance, but did not normalize the mean of S_{it} to zero as it is of interest to ask whether the individuals are on average attracted to or repulsed from conspecifics. For simplicity, we assume the linear regression model

$$S_{it} = \mu + \mu_i^H + \mu_{it}^T + (\beta + \beta_i^H + \beta_{it}^T)P_{it} + \varepsilon_{it},$$

where the intercept μ estimates the population mean level of sociability and the slope β the effect of the proximity of conspecifics on sociability. The model has two random components which relate to population heterogeneity (superscript H for random variation among individuals) and to temporal heterogeneity (superscript T for random variation within individuals over time). Both of these random components are assumed to follow bivariate normal distributions with covariance structures $\text{Cov}(a_i^H, b_{i'}^H) = \delta_{ii'} \Sigma_{ab}^H$ and $\text{Cov}(a_{it}^T, b_{it'}^T) = \delta_{ii'} \Sigma_{ab}^T f(t, t', T)$, where $a, b \in \{\mu, \beta\}$. Here Σ^H and Σ^T are 2 x 2 variance-covariance matrices, and $\delta_{ii'}$ is Kronecker's delta specifying that the effects are independent among the individuals. We model temporal autocorrelation assuming exponential decay at temporal scale T , $f(t, t', T) = \exp(-|t - t'|/T)$, and assume that the residual variation ε_{it} follows normal distribution with zero mean and variance σ^2 .

We used Bayesian inference to estimate model parameters. The choices of priors for model parameters $(\mu, \beta, \Sigma^H, \Sigma^T, T \text{ and } \sigma^2)$, as well as a description of the Markov chain Monte Carlo (MCMC) algorithm used to sample the posterior distribution are described in Appendix A. We examined how the squared response in the fitted model $V = E[S^2]$ is split up into the four components of the model (1): variation V_P related to population mean, variation V_H related to population heterogeneity, variation V_T related to temporal heterogeneity, and residual variation V_R . Expanding S^2 and noting that the covariate P is normalized to zero mean and unit variance, these can be written as $V_P = \mu^2 + \beta^2$, $V_H = \Sigma_{\mu\mu}^H + \Sigma_{\beta\beta}^H$, $V_T = \Sigma_{\mu\mu}^T + \Sigma_{\beta\beta}^T$ and $V_R = \sigma^2$.

GENERATION OF SIMULATED DATA

The simulations were conducted in a square with area 10 km x 10 km. We assumed the average density to be one individual per km² and the simulation was initialized by randomizing the 100 individuals according to complete spatial randomness (CSR). With CSR and density 1/km², the mean distance to the nearest neighbour is 500 meters, with 95% quantile ranging from 90 meters to 1.1 km (Kingman 1993). To mimic infinite space, we assumed toroidal boundary conditions. We simulated the chosen process (see below) with day as a time step, first removing a transient of 100 days, and then collecting data for $T = 100$ days. In all models, individuals moved with constant steps of 100 m each day. The decision of where to move next (day $t+1$) was based on information about the current location of the other individuals (day t).

Data were simulated according to the following six scenarios. In Scenario 0 (independent movements), the direction of the next step is random. In Scenario 1 (preferred distance to the nearest conspecific), the focal individual prefers to have a distance of 250 meters to the nearest conspecific, thus creating a tendency to form aggregations with circa four-fold density relative to the case of CSR.

1 The individual samples $k = 5$ potential locations at random directions from the present location, and
2 chooses the one which is closest to its preference. Note that higher values of k would lead to more
3 deterministic movements towards the preferred state, whereas $k = 1$ would result in a simple random
4 walk; k therefore captures the completeness of the individuals knowledge of its surroundings. In
5 Scenario 2 (preferred local density of conspecifics), the individual prefers to have 12 other individuals
6 in a circle of 1 km radius from its location, corresponding again to ca. four-fold density compared to
7 the mean density. As in Scenario 1, the next location is selected as the most preferred one among $k = 5$
8 randomly selected locations. In Scenario 3 (heterogeneous population), half of the individuals are
9 socially independent and half socially dependent, the rules being like in Scenarios 0 and 1, respectively.
10 In Scenario 4 (behavioural switching) we consider that individual decisions regarding conspecifics
11 might change over time depending on different ecological circumstances (e.g. predation risk, resource
12 depletion), and thus each individual has two possible behavioural modes, socially independent and
13 socially dependent, the rules being as in Scenarios 0 and 1, respectively. The initial condition is as in
14 Scenario 3, after which at each time step, each individual switches its mode with probability $p = 0.1$. In
15 Scenario 5 (confounding factors), we simulated an underlying resource distribution that changes in time
16 and space. Resources are modelled as points, which are initially randomly distributed over the domain
17 with density $\rho = 0.1$ per unit area. At each time step each resource point is displaced (i.e. disappears
18 and appears in a new random location) with probability $q = 0.1$. Each individual samples $k = 5$
19 locations, and selects the one which is closest to the nearest resource point, independent of the presence
20 of conspecifics. Under this scenario, individual movements may appear to be dependent on the
21 behaviour of conspecifics, though there are no direct social interactions.

22 We first asked whether the statistical models described in the previous section are structurally
23 adequate for capturing the mechanisms behind the individual-based simulation models described in this

1 section. Here we assumed the availability of much data (100 steps observed for 100 individuals) and
2 applied the proximity measure that corresponded to each of the scenarios (density at spatial scale of 1
3 km for Scenario 2, nearest neighbour measure for the other scenarios). Second, to ask if the model
4 could correctly identify the spatial scale at which the decisions were made in the simulations, we fitted
5 the data on Scenarios 1 and 2 with both proximity measures (for density with spatial scales $d =$
6 0.25, 0.5, 1.0, 2.0, 4.0 km). Finally, we assessed the performance of the statistical models to work with
7 incomplete data by modifying the Scenarios 3 and Scenario 4 as follows. To examine the effect of
8 missing individuals, we assumed that we had data only on 25 or 50 out of the 100 individuals. To
9 examine the effect of the sampling interval, we sub-sampled the data two- or fourfold, resulting into 50
10 or 25 observed movement steps during the study period of $T = 100$ days.

11

12 EMPIRICAL DATA ON EAGLE OWLS

13 We focus here on the wandering phase of dispersal, i.e. when individuals explore different areas for a
14 variable time period before settling in a new area (Delgado *et al.* 2010), of 55 juvenile eagle owls (36
15 males and 19 females) radiotagged in south-western Spain during the years 2003-2007. Owls were
16 located at their diurnal roosts ($n = 766$; 538 locations for males and 228 locations for females), the data
17 including on average 12.8 (sd = 7.1) observations for each individual, sampled on average at 10.5 day
18 (sd = 11.6) intervals. We used a slightly modified version of the statistical modelling framework
19 described above to ask if and how owls responded to conspecifics during dispersal. First, when
20 computing P^{null} , we accounted for variation in step lengths by fitting a linear model $\log(\text{step length}) \sim$
21 $\log(\text{duration of sampling interval})$ separately for each owl, and then used these models to sample step
22 lengths in the null model. Thus, the response variable S_{it} here reflects not only whether the owl
23 avoided or was attracted to other owls in terms of the direction, but also in terms of the distance it

1 moved. Second, we accounted for a possible preference for habitat by accepting in the null set only
2 locations which consisted of the same habitat as the location the individual moved to in reality (see
3 Delgado *et al.* 2010 for habitat classification). As location data for all individuals were not obtained
4 simultaneously, the locations of the other owls were typically not known exactly at the time when the
5 observation of the focal owl was made. As a simple solution to this problem, we used linear
6 interpolation between observed locations to estimate the location of each owl at any time.

7 We assumed that individual movement response to the social environment may depend on the state
8 of the individual, for which we considered the three variables of (1) sex, (2) body condition index
9 (BCI) and (3) haematocrit index. Higher values of parameters 2 and 3 correspond to better individual
10 condition (for more details on the empirical study and the measurement of these two indices, as well for
11 earlier results, see Delgado *et al.* 2010). We modelled the effect of the individual as $\mu_i^H = \sum_k x_{ki} \gamma_k +$
12 $\mu_i^{RH}, \beta_i^H = \sum_k x_{ki} \vartheta_k + \beta_i^{RH}$, where x_{ki} is the covariate of type $k = 1, 2, 3$ measured for individual i , and
13 parameters γ_k and ϑ_k measure the effect of covariates on the behavioural rule, i.e. to the overall level
14 of sociability and its dependence on conspecifics. The residual components of variation μ_i^{RH} and β_i^{RH}
15 were modeled as μ_i^H and β_i^H in the original model. To explore whether individuals were responding to
16 overall conspecific density or specifically to the density of the same or opposite sex, we fitted the
17 models so that proximity (and thus also sociability) was based solely on females or males.

18 The statistical models were implemented in Mathematica (source code and binaries can be found in
19 supplements).

20 21 **Results**

1 In general, the partitioning of variance in the fitted statistical models was consistent with the simulated
2 behaviour of the individual based models. When movements were not affected by conspecifics
3 (Scenario 0; Fig. 2a), almost all model variance was assigned to the residual. For Scenarios 1 and 2
4 (Figs. 2b and 2c, respectively), almost all of the non-residual variance was assigned to component V_P ,
5 indicating that the individuals were similarly affected by their social environment. For the case of a
6 heterogeneous population (Scenario 3; Fig. 2d), a substantial amount of variation was assigned to
7 component V_H measuring variation among individuals. For the case of behavioural switching (Scenario
8 4; Fig. 2e) a considerable proportion of variation was assigned to component V_T which measures
9 temporal heterogeneity within the individuals. In this scenario, the temporal scale of the autocorrelation
10 reflected the frequency with which the individuals switched their behaviour. In the case where the
11 individuals tracked the changing distribution of resource points (Scenario 5; Fig. 2f), the fitted
12 statistical model resembled that of behavioural switching, though with a low scale of temporal
13 autocorrelation. Figure 2 shows that for both Scenarios 1 and 2, the best model fit in terms of lowest
14 residual variance was obtained for the statistical model with nearest neighbour proximity. We note that
15 for Scenario 2 a large amount of variation in this model was assigned to temporal heterogeneity V_T with
16 temporal scale T very close to zero; in this case the variance component V_T is equivalent to residual
17 variation (see Appendix A). Therefore, for Scenario 2 the best fitting statistical model was the one
18 where proximity is based on individual density within 1 km around the focal individual, i.e. the model
19 that corresponds with the simulated scenario.

20 These results indicate that the statistical models can correctly describe the underlying movement
21 behaviour when abundant and high quality data are available. As expected, the ability of these models
22 to disentangle the variance components decreases both with the number of missing individuals and with
23 the length of the sampling interval (Fig. 2d,e). The identification of population heterogeneity is more

sensitive to the number of sampled individuals than to the sampling interval, whereas data for many steps per individual are required to identify behavioural switching. Repeating the analyses with measure S^{t_2} lead to similar results though with a slightly worse model fit than with measure S^{t_1} (Appendix B).

EMPIRICAL DATA ON EAGLE OWLS

In the case of eagle owls, the proximity measure leading to the smallest residual variation was the nearest-neighbour measure (Fig 2g, Appendix C). In this model, the variance was partitioned as $V_P = 0.14$ (95% highest posterior density interval 0.06-0.25), $V_H = 0.18$ (0.09-0.32), $V_T = 0.43$ (0.26-0.60) and $V_R = 0.23$ (0.11-0.40), indicating a substantial amount of variability both among individuals and within individuals, the latter with temporal scale $T = 5.9$ (3.4-10.5) days. Owls were generally attracted to conspecifics, the strength of the attraction decreasing with increasing proximity (Fig. 1c). Females were more attracted to conspecifics than males, and both males and females were more attracted towards females than males (Fig. 1c). Owls in good condition were generally less attracted to conspecifics than owls in poor condition, but this statement obtained only limited statistical support (Appendix C).

Discussion

We have presented a statistical framework for inferring the influence of conspecifics on movement behaviour and illustrated its use both with simulated data and with a case study on eagle owl dispersal. Unlike much of the earlier work on collective movement behaviour (e.g. Couzin *et al.* 2002; Guttal & Couzin 2010), our approach is not based on specific assumptions about the underlying mechanisms, but is purely statistical. Our framework is related to previous approaches developed to infer drivers of

1 animal space use patterns such as Step Selection Functions (SSF; Fortin *et al.* 2005) in the sense that
2 the locations of other individuals can be considered as a spatial factor that influences the individual
3 movement decisions. In our framework, we have included all other aspects normally included in SSF
4 (i.e. those related to habitats, movement modes, etc.) in the null model and thus separated them from
5 the influence of conspecifics. The first reason for doing so is that this allows making the conceptual
6 framework related to proximity and sociability (Fig. 1) explicit. The second and somewhat related
7 reason is that this allows for a more straightforward interpretation of the statistical model. The
8 advantage of the present approach is in its generality: it can be applied to almost any movement data on
9 interacting entities. However, a more mechanistic approach tailored to the specifics of a given case
10 study is likely to yield more detailed information. Thus, we see our approach as a first step in the
11 analysis of movement data potentially influenced by conspecific behaviour.

12 We have developed our approach in the context of hierarchical Bayesian models, which are
13 becoming increasingly popular in many studies of ecological and evolutionary processes (Cressie *et al.*
14 2009), largely due to their flexibility in the accommodation of multiple levels of dependency, modelled
15 through the effects of covariates and correlation structures (Gillies *et al.* 2006). We have selected the
16 use of Bayesian statistics mainly for the reasons of computational feasibility. However, as our model is
17 simply a regression model with fixed and random effects (Dingemanse *et al.* 2012), it could be fitted to
18 the data using also frequentist techniques. In either case, the practitioner needs to carefully check (like
19 in any modelling exercise) that the assumptions of the model (such as normality and homoscedasticity
20 of residual variance) are met, though our results based on simulated data turned out to be robust even
21 though these assumptions were not necessarily always met.

22 As demonstrated by the application of the model to data generated through the simulation of
23 different scenarios, the estimated variance components can be related in a biologically meaningful way

1 to the actual mechanisms behind socially informed movement decisions. The main reason for focusing
2 here on variance partitioning is that we found it to be the simplest and most robust way to quantify the
3 importance of each behavioural component. In practice, if a practitioner fits our statistical model to
4 their data, then the question of "are the individuals responding to each other and in which way?" can be
5 addressed by looking at how much of the total variation is assigned to each component of the statistical
6 model. In the eagle owl example, we found that dispersing individuals were attracted to others and that
7 their movement rules varied not only among individuals, partly explained by measured variation in
8 their internal states, but also within individuals.

9   There are two important considerations to be made when applying the framework developed here.
10 First, the temporal resolution at which individuals make their movement decisions is usually unknown
11 *a priori*. If individuals are sampled at very coarse temporal resolution, the statistical power for
12 detecting interaction effects may be limited (Dingemanse *et al.* 2012), as we illustrated by sub-
13 sampling the simulated data. On the other hand, if individuals are sampled at a temporal resolution
14 much higher than the interval at which movement decisions are made, then consecutive data points
15 become statistically dependent on each other, potentially leading to a false impression of behavioural
16 switching. This may be partially mitigated by having a more sophisticated null model which includes
17 autocorrelation in the movement.

18   Second, the fact that individuals live or move in a group does not necessarily imply that they
19 interact socially (Bode *et al.* 2012). As demonstrated by the simulated Scenario 5, if individuals are
20 attracted to the same resources (e.g. food or shelter), their movements may appear to be socially
21 dependent even if this is not the case. It is worth noting the fact that, in this scenario, the scale of the
22 behavioural responses of individuals is likely to be affected by the spatial aggregation and the temporal
23 predictability of resources that we considered (e.g. Fauchald & Tveraa 2006). Again, in theory, this

1 problem can be mitigated by using a more sophisticated null model, in which all the non-social factors
2 influencing movements are accounted for. This is by no means a trivial task, as many of these factors
3 are likely to be unknown, and even the influences of the known factors can be hard to quantify reliably.
4 Therefore, when applying our method to any real movement data, the practitioner needs to carefully
5 think about the selection and parameterization of the null model, as in theory the null model should
6 incorporate all other factors influencing movement decision excepting the influence of conspecifics.
7 While it is clear that the construction of a perfect null model is difficult if not possible, any other
8 approach attempting to pinpoint the influence of conspecifics on movement decisions from non-
9 manipulative observational data necessarily has the same shortcoming of the results being sensitive to
10 the assumptions made on other factors influencing movement behaviour. Whatever the object of
11 analysis, the choice of the temporal scale at which individuals are to be sampled, the construction of the
12 appropriate null model, and the interpretation of the statistical results all require a good understanding
13 of the ecology and life history of the study species.

14 Our simulation experiments illustrate the ability of the proposed statistical framework in
15 pinpointing different kinds of underlying mechanisms behind movement behaviour, as well as how its
16 statistical power may depend on some issues common in real data (such as missing individuals).
17 However, it was not feasible in this study to exhaustively address all possible ways in which different
18 confounding factors (e.g. spatial and temporal changes of habitat quality across the landscape) may
19 lead to similar movement patterns than those generated by interaction individuals. When applying the
20 method presented here, we encourage the practitioner to perform simulations that emulate the potential
21 confounding factors of the system under study.

22 As was the case with the owl study, spatial positions are often not taken simultaneously for all
23 individuals. Linear interpolation, or any other method to estimate the missing locations, will add noise

to the measures of proximity and sociability, and thus it is likely to reduce the ability of the statistical framework to detect non-random behaviour, making the results conservative.

Our study illustrates the general applicability of hierarchical Bayesian models in ecology (Clark 2005). The framework we have developed applies in principle to movement data on any kinds or particles, from organelles within cells to GPS tracking data on large animals. The main limitation of the application of the methods is the need of high resolution movement data acquired simultaneously for a sufficiently large number of individuals. While such data are currently still rare, we expect that the ongoing revolution in tracking technology (Cagnacci *et al.* 2010) will result in a high flow of such data in the future.

Acknowledgements

The work was funded by one research project of the Spanish Ministry of Economy and Competitiveness (CGL2012-33240; with FEDER co-financing) and the Junta de Andalucía (Excellence Project, RNM-5090). MD (#14037) and OO (#250444) were financially supported by the Academy of Finland and OO by the European Research Council (ERC StG 205905). JMM was supported by CONICET (PIP 114 - 200801 – 00276).

References

- Bode, N.W.F., Franks, D.W., Wood, A.J., Piercy, J.J.B., Croft, D.P. & Codling, E.A. (2012) Distinguishing Social from Nonsocial Navigation in Moving Animal Groups. *American Naturalist*, **179**, 621-632.
- Cagnacci, F., Boitani, L., Powell, R.A. & Boyce, M.S. (2010) Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 2157-2162.
- Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters*, **8**, 2-14.

- Couzin, I.D., Ioannou, C.C., Demirel, G., Gross, T., Torney, C.J., Hartnett, A., Conradt, L., Levin, S.A. & Leonard, N.E. (2011) Uninformed individuals promote democratic consensus in animal groups. *Science* 334, **334**, 1578-1580.
- Couzin, I.D., Krause, J., James, R., Ruxton, G.D. & Franks, N.R. (2002) Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, **218**, 1-11.
- Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553-570.
- Delgado, M.D., Penteriani, V., Revilla, E. & Nams, V.O. (2010) The effect of phenotypic traits and external cues on natal dispersal movements. *Journal of Animal Ecology*, **79**, 620-632.
- Dingemanse, N.J., Bouwman, K.M., van de Pol, M., van Overveld, T., Patrick, S.C., Matthysen, E. & Quinn, J.L. (2012) Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *Journal of Animal Ecology*, **81**, 116-126.
- Eriksson, A., Jacobi, M.N., Nystrom, J. & Tunstrom, K. (2010) Determining interaction rules in animal swarms. *Behavioral Ecology*, **21**, 1106-1111.
- Fauchald, P. & Tveraa, T. (2006) Hierarchical patch dynamics and animal movement pattern. *Oecologia*, **149**, 383-395.
- Fletcher, R.J. (2006) Emergent properties of conspecific attraction in fragmented landscapes. *American Naturalist*, **168**, 207-219.
- Fletcher, R.J. (2007) Species interactions and population density mediate the use of social cues for habitat selection. *Journal of Animal Ecology*, **76**, 598-606.
- Fortin, D., Beyer, H.L., Boyce, M.S., Smith, D.W., Duchesne, T. & Mao, J.S. (2005) Wolves influence elk movements: Behavior shapes a trophic cascade in Yellowstone National Park. *Ecology*, **86**, 1320-1330.
- Gillies, C.S., Hebblewhite, M., Nielsen, S.E., Krawchuk, M.A., Aldridge, C.L., Frair, J.L., Saher, D.J., Stevens, C.E. & Jerde, C.L. (2006) Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology*, **75**, 887-898.
- Guttal, V. & Couzin, I.D. (2010) Social interactions, information use, and the evolution of collective migration. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16172-16177.
- Hemelrijk, C.K., Hildenbrandt, H., Reinders, J. & Stamhuis, E.J. (2008) Emergence of Oblong School Shape: Models and Empirical Data of Fish. *Ethology*, **116**, 1099-1112.
- Herbert-Read, J.E., Perna, A., Mann, R.P., Timothy M. Schaerf, Sumpter, D.J.T. & Ward, A.J.W. (2011) Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences*, **108**, 18726-18731.
- Katz, Y., Tunström, 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*, **108**, 18720-18725.
- Kawaguchi, L.G., Ohashi, K. & Toquenaga, Y. (2006) Do bumble bees save time when choosing novel flowers by following conspecifics? *Functional Ecology*, **20**, 239-244.
- Kingman, J.F.C. (1993) Poisson processes. *Clarendon Press, Oxford*.
- Kuussaari, M., Saccheri, I., Camara, M. & Hanski, I. (1998) Allee effect and population dynamics in the Glanville fritillary butterfly. *Oikos*, **82**, 384-392.
- Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E. & Fryxell, J.M. (2004) Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology*, **85**, 2436-2445.

- 1 Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. (2008) State-space
2 models of individual animal movement. *Trends in Ecology & Evolution*, **23**, 87-94.
3 Serrano, D., Tella, J.L., Donazar, J.A. & Pomarol, M. (2003) Social and individual features affecting
4 natal dispersal in the colonial Lesser Kestrel. *Ecology*, **84**, 3044-3054.

5

6 **Supplemental Material**

7 **Appendix A.** Formal description of the hierarchical Bayesian model.

8 **Appendix B.** Performance of the statistical framework with measure S^{t_2} .

9 **Appendix C.** Parameter estimates for the eagle owl case study.

10 **Supplements.** Mathematica code and eagle owl data for fitting the hierarchical Bayesian model.

Figure legends

Figure 1. Panel (a) shows the relationship between current proximity (P) to other individuals and the expected sociability (S) for three hypothetical scenarios: (i) individual attracted to other individuals irrespective of its current proximity (blue line), (ii) individual attracted to conspecifics at low but repulsed at high density (black and red lines), and (iii) individual moving independently of conspecifics (dashed-line). Panel (b) illustrates how sociability can be measured from movement data. Large dots represent the focal individual at its initial position (black dot) and after a movement step (red dot). Small dots show the locations of six other individuals. After the movement step, the focal individual is close (39 m) to where one of the conspecifics was initially. The random expectation, averaged over locations within the blue circle, is 69 m, thus the focal individual showed attraction ($S^{t_1}=30$ m). Panel (c) shows the median estimates of the functional response for the owl case study. The red and blue lines correspond to the focal owl being a female or male, respectively. The line style indicates how proximity to other individuals is measured: all individuals (continuous line), females only (dashed line), or males only (dotted line).

Figure 2. Performance of the statistical framework for inferring movement behaviour from simulated (a-f) and eagle-owl (g) data, including the ability of the models to pinpoint the spatial scale at which the individuals assess proximity (for Scenarios 1 and 2), and the effect of sample size in terms of number of tracked individuals and the sampling interval (for Scenarios 3 and 4). In all panels, bars show averages (over replicates and posteriors) of the four variance components. The smaller (bold) error bars show parameter uncertainty measured as 95% quantile in the posterior distribution (averaged over replicates). The larger (light) error bars show additionally variation among replicates (minimal and maximal values among the 10 replicates). NN: nearest neighbour proximity measure. For parameter values, see main text.

Fig. 1

1 **Fig. 2**

2

3