MODELS AND MEASURES OF ANIMAL AGGREGATION AND DISPERSAL

MARK BROOM[1], IGOR V. EROVENKO[2], JONATHAN T. ROWELL[2], AND JAN RYCHTÁŘ*[3]

Abstract. The dispersal of individuals within an animal population will depend upon local properties intrinsic to the environment that differentiate superior from inferior regions as well as properties of the population. Competing concerns can either draw conspecifics together in aggregation, such as collective defence against predators, or promote dispersal that minimizes local densities, for instance to reduce competition for food. In this paper we consider a range of models of non-independent movement. We include established models, such as the ideal free distribution, but also develop novel models, such as the wheel. We also develop several ways to combine different models to create a flexible model of addressing a variety of dispersal mechanisms. We further devise novel measures of movement coordination and show how to generate a population movement that achieves appropriate values of the measure specified. We find the value of these measures for each of the core models described, as well as discuss their use, and potential limitations, in discerning the underlying movement mechanisms. The movement framework that we develop is both of interest as a stand-alone process to explore movement, but also able to generate a variety of movement patterns that can be embedded into wider evolutionary models where movement is not the only consideration.

1. Introduction

Most organisms, especially higher order animals, experience a non-sessile phase during their life history. These periods of movement may represent postnatal dispersal, active foraging to obtain resources or mating opportunities, nomadism, or seasonal migrations between distant breeding and non-breeding sites (Dingle and Drake, 2007). The identification of underlying mechanisms

---

[1] Department of Mathematics, City, University of London, Northampton Square, London, EC1V 0HB, UK

[2] Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC 27402, USA

[3] Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, VA 23284-2014, USA

Key words and phrases. patch selection; Ideal Free Distribution; habitat choice; group movement coordination; animal movement; urn models.

* Corresponding author: rychtarj@vcu.edu.

for the spatial behavior of properties has been the subject of considerable inquiry and musing for decades (Travis et al., 1999; Ford and Swearer, 2013). Social interactions and communication may influence migration of groups of animals as well as other collective actions particularly when there is evidence of a high degree of synchrony among individuals (van Noordwijk et al., 2006; Kappeler, 2011; Petit and Bon, 2010; Pyritz et al., 2011). Here we promote a general framework of movement that incorporates established and novel concepts of dispersal and patch selection that may be positively or negatively dependent upon the presence of conspecifics. This framework is robust and may be incorporated into general evolutionary modelling approaches, in particular that of Broom and Rychtář (2012), including systems where movement is only one factor influencing fitness. In this paper we will specifically focus on the modelling of movement.

Migration is present throughout the taxa (Dingle and Drake, 2007); however, not all relevant species exhibit the traditional round-trip or "two worlds" movement (Greenberg and Marra, 2005) that is characteristic of large-scale synchronized migration of e.g. birds (Dingle, 2006). In the literature, the term movement or migration has been applied to different situations including post-natal dispersal and other one-phase relocations of organisms (Greenwood and Harvey, 1982; Bullock et al., 2002), ranging (Kennedy, 1966; Taylor, 1986; Dingle, 2014), or nomadism (Dingle and Drake, 2007). For our purposes in this paper, we will assume a minimal definition of movement as the selection of a next place of residence whether the location is held permanently (settlement) or in transience (migratory stopovers, ephemeral ponds). Aggregation during migration has been observed in neritid snails (Schneider and Frost, 1986), mayflies (Hayden and Clifford, 1974), insect swarms (Buhl et al., 2006), spiny lobsters (Herrnkind and Cummings, 1964), fish schools (Ballerini et al., 2008; Parrish, 1989, 1999; Parrish et al., 2002), bird flocks (Emlen, 1952; Ballerini et al., 2008), and primates (Pyritz et al., 2011). Additionally some birds forage collectively in conspecfic groups (Beauchamp, 1998, 2002; Krause and Ruxton, 2002), and many predator mammals hunt collectively as well (Würsig, 1986; Mech, 1981; Kruuk, 1972; Stander, 1992). The social cohesion of animal groups is subject to density dependent mechanisms that may be in tension with some promoting dispersive behavior in populations and others facilitating aggregation (Ford and Swearer, 2013; Matthysen, 2005; Cressman and Garay, 2011; Parrish, 1999). As an example, ducks have been found to assess the local conditions of sites and are even capable of adaptively adjusting their time in residence at a stopover in response to conditions (O'Neal et al., 2012), yet they will initially land in close proximity to one another for safety in novel territory. Resource limitation is perhaps the most important driver of population regulation that promotes dispersal (Murdoch, 1994; Turchin, 1999); however, for species subject to the social fence hypothesis (Hestbeck, 1982), high densities can inhibit dispersal and keep groups together. Moreover, conspecific attraction (Danielson and Gaines, 1987; Stamps, 1991), Allee effects (Kuussaari et al., 1998), and safety-in-numbers effects (Beauchamp, 2008) may also reinforce congregation within animal populations. The resulting benefits of social aggregations include anti-predator behavior (Ioannou and Krause, 2008; Pitcher et al., 1982; Cressman and Garay, 2011), increased survival (White et al., 2010; Ford and Swearer, 2013), social foraging efficiency (Pitcher et al., 1982; Felleman, 1986), cues for the availability of resources or mates (Pitcher et al., 1982; Matthysen, 2005; Guttal and Couzin, 2010; Kim et al., 2009), and, crucially for migratory animals, energetic efficiencies such as drag reduction (Ward et al., 2001). These features are of particular importance for colonial breeders who may associate in obligate fashion (Kim et al., 2009).

The dispersal of a population over a territory, either as a single event or as a sequence of moves, has been considered in a number of models. The classical models of animal distribution over a territory consider the distribution of animals over a number of patches of a resource, such as food or mates. Probably the most well known model is the ideal free distribution (IFD) of Fretwell and Lucas (1969). Here individuals choose from a variety of food patches to maximize their intake, and this leads to an effectively deterministic division of forages with all receiving the same reward. For an explicitly game-theoretical interpretation of the IFD see Cressman and Křivan (2006); Křivan et al. (2008); Cressman et al. (2004). For a model including migration see Mariani et al. (2016) and for one including harvesting see Křivan and Jana (2015).

There are both deterministic and stochastic models which build upon the IFD. The "perceptual limit" model was developed in Abrahams (1986), where animals moved randomly between patches if the differences between the reward on two patches was below a given level (see also Collins et al., 2002; Street et al., 2018). The IFD was refined to allow for unequal or non-identical competitors in Sutherland and Parker (1992). There may also be non-IFD movements caused by events not linked to the resource value, for example the arrival of a predator (see Hugie and Grand, 1998; Ruxton and Humphries, 2003). Stochastic models of unequal competitors were considered in Houston and McNamara (1988) and Yates and Broom (2005).

In Broom and Rychtář (2012) a general framework for the movement of, and interactions between, members of a population was considered. Most models using this framework up until now, e.g. Broom et al. (2015), Pattni et al. (2017) have used the "fully independent" model where each individual has its own probability distribution over the patches, independent of other individuals and past movements. Recent works Pattni et al. (2018) and Erovenko et al. (2019) have considered movement dependent on the most recent history; but still here, conditional on the history, all individuals move independently.

The models used in this paper will serve two related purposes; firstly to represent particular movement mechanisms, which lead to a given distribution of individuals over patches, and secondly to model movement distributions with certain aggregational properties. We consider a series of models where individuals choose their location sequentially, based upon a utility function approach. These include purely deterministic models where the best utility is selected and stochastic models where the probability of selecting a location is a function of the utilities. We also consider a model of simultaneous allocation which does not represent any specific movement mechanism, but is useful in achieving desirable statistical properties of the population (for instance we may wish to model a population with a given distribution and mean pairwise correlation over the patches). We consider some interesting special models, and general ways of combining models to again achieve desirable statistical outcomes.

The paper is organized as follows. In Section 2 we consider sequential models of patch selection. These models are classical (such as the Ideal Free Distribution in Section 2.1.3), although they have not always been applied to modeling animal movements (such as probabilistic models in Section 2.2). In Section 3 we consider completely novel models and approaches to modeling animal movement. We also develop a way to compare different models. In Section 4 we introduce and study novel measures of movement coordination. Finally, in Section 5 we provide and discuss biological examples for our models.

2. SEQUENTIAL MODELS OF CORRELATED PATCH SELECTION

In this section we introduce several possible models of how N individuals $I_1, I_2, \ldots, I_N$ can distribute themselves over M patches $P_1, P_2, \ldots, P_M$. We will usually assume that the individuals are identical and have the same movement distributions, i.e. the probability of an individual I_i moving to a patch P_m will depend only on the patch (and nothing else). We shall use sequential movement mechanisms where individuals use information on the distribution of earlier-arriving animals, but then identical distributions over all individuals can be obtained by allocating the order sequence of individuals at random. These mechanisms are “realistic”, in the sense that it is possible for individuals to use this kind of mechanism in a sequential process. Also, these mechanisms are based on known models, although as far as we know, the models were not always applied to or developed for animal movements. For each such model from Sections 2.1–2.3, we consider an *a priori* probability distribution $\{a_m\}$, i.e. without considering any other factors, an individual would select a patch P_m

with a probability a_m. This distribution reflects a natural inclination in location preference during immigration into the area irrespective of present conditions. For example, site fidelity in migratory species exhibits this phenomenon. The *a priori* distribution could also represent the accessibility of sites upon first arrival. The actual probability distribution of an individual's patch selection may evolve as the landscape becomes increasingly populated.

We note here that as all individuals are identical, the precise ordering of individuals is irrelevant. If there was some distinction between the individuals, the order might be important, and we might need to consider a probability distribution over the possible orders. As mentioned above, we would then randomly pick a permutation π of N numbers and then place the individuals into the patches sequentially in the order dictated by the permutation, i.e. we first place an individual $I_{\pi(1)}$, then individual $I_{\pi(2)}$, etc. We could also represent simultaneous movement by selecting each of the $N!$ permutations with equal probability.

In general, when $n - 1 \in \{0, 1, \dots, N - 1\}$ individuals are already placed, the n^{th} individual $I_{\pi(n)}$ will move to a patch P_m with a probability that depends on the placements of all of the previous individuals $I_{\pi(1)}, I_{\pi(2)}, \dots, I_{\pi(n-1)}$ (but it does not depend on which individuals are actually placed there since all individuals are assumed identical). After the placements, we calculate p_m, the *a posteriori* probability that an individual goes to patch P_m as the expected value over all possible permutations (and consequently, does not depend on the individual).

As we will see below, in some cases (such as follow the majority, random, Polya urn), the *a priori* probabilities will agree with the *a posteriori* probabilities and we will call such procedures *faithful*. However, in other cases (such as competitor avoidance, peak overload and IFD), the *a priori* and *a posteriori* probabilities do not match (except for in some very specific cases).

2.1. **Deterministic selection models.**

Here we assume that there is a set of utility functions $\{U_m\}$ based upon various patch characteristics. The individuals maximize their utility and so the n^{th} individual is placed to P_m if U_m is the unique maximum utility. In circumstances where maximal utility is achieved at multiple locations, the individual selects its location randomly from among those locations $\{P_{m_k}\}$ with probability proportional to the *a priori* probabilities $\{a_{m_k}\}$.

For simplicity, we assume that the utility function for patch P_m is potentially dependent only upon an intrinsic non-negative patch value V_m and the patch's current occupancy Y_m. In particular, the utility value is independent of conditions in other patches within the landscape. Broom and Rychtář (2018) applied this local aggregation assumption to structured population models. This assumption is reasonable if there is not some direct influence of one patch on another, such as if our individuals have a highly mobile predator who can observe more than one patch, so that its choice to hunt on one patch depends upon what it observes on another. Additionally, we further assume that utility increases with the value of the patch. We will distinguish several cases based upon how the utility depends on the local population size.

2.1.1. *Utility is independent of patch occupancy.*

Here the utility is simply $U_m = V_m$, the intrinsic value of the patch P_m. In this case, the individuals will all aggregate at a patch of the highest value. We will refer to this model as *peak overload.*

If $\overline{M} = \{m_1, \dots, m_K\}$ is the set of indices for those patches which support maximal utility, then for all $m \in M$, the expected number of individuals at patch P_m is given by

$$E[X_m] = \begin{cases} \dfrac{a_m}{\sum\limits_{m_k \in \overline{M}} a_{m_k}} N, \text{if } m \in \overline{M}, \\ 0, \text{otherwise.} \end{cases} \tag{1}$$

In the non-generic case where one patch is of higher value than the others, all individuals will go to this patch, i.e. the expected number of individuals will be N on this patch and 0 elsewhere, with a variance of 0 on all patches. With K equal patches, we can simply ignore all other patches (which will be empty) and otherwise this case will be the same as for the random allocation method, which we see in Section 2.2.1, over these K patches.

Biologically, settlement without respect to population pressures is appropriate when locations provide non-depletable resources or when they have an ideal environmental feature. Examples of this include temperature, water pH, salinity, sunlight exposure, etc. These conditions may make locations suitable for a nursery or nesting site. Furthermore, as noted in Broom and Rychtář (2018), the concept of "place" or "patch" can be quite general. Thus if "patches" represent migration routes or layover territories, one route may be better than others for causes beyond the presence of conspecifics using the same transient location or route.

We see here that all individuals end up on the best patch, if that is unique, and they can be thought of as forming a "herd". We note, however, that the herd is formed simply because all individuals choose the best patch independently, and not through any wish of individuals to be with others. We will use the term *herding* for any procedure that will result in all individuals aggregating in one patch when this is caused by some inherent desire to be with others. We will see examples of herding models below in Section 2.1.2 and Section 2.3.

2.1.2. *Utility positively correlates with occupancy.* When the utility functions increase with local population size, individuals prefer to stay in large conspecific groups or herds. In principle we could consider any increasing function. However, we shall only consider the simplest example utility function, namely

$$U_m = Y_m + 1, \tag{2}$$

where Y_m is the number of current occupants at the patch P_m. We will refer to this model as *follow the majority* model.

In accordance with our methodology, the first individual is placed randomly using the *a priori* probability distribution $\{a_m\}$ as all sites are of maximal utility. All subsequent individuals will then move to the same patch as the initial settler. Consequently, this is an example of a herding procedure.

The expected number of individuals at patch P_m is given by $E[X_m] = a_m N$, thus $p_m = a_m$ and the procedure is faithful. Moreover

$$\begin{aligned} \mathrm{Var}(X_m) = E[X_m - E(X_m)]^2 &= p_m(N - Np_m)^2 + (0 - Np_m)^2(1 - p_m) \qquad (3) \\ &= N^2 p_m(1 - p_m). \qquad (4) \end{aligned}$$

2.1.3. *Utility negatively correlates with occupancy.* Conspecific pressures commonly reduce the utility of a location as limited resources must be shared among all of those present or as internal conflicts arise within the local population, and, as above, any decreasing function could be considered as a utility function. The *ideal free distribution* is a natural example of populations dispersing to balance these pressures against the value of various locations. In a discrete sequential population, the utility functions are given by

$$U_m = \frac{V_m}{Y_m + 1}, \tag{5}$$

where V_m is the intrinsic value of the patch P_m and Y_m is the number of its current occupants. Here the utility is calculating the post-settlement fitness of the individual rather than the current per capita utilization of the patch. The sequence of settlement follows a predictable outcome, but in those instances where multiple patches predict maximal utility, the individual randomly selects from those patches in proportion to their *a priori* fitness (see above in Section 2.1.1).

As the total population size N grows large, the expected proportion of individuals on patch P_m approaches

$$\frac{E[X_m]}{N} = \frac{V_m}{\sum_{k=1}^{M} V_k}, \tag{6}$$

i.e. we have an *a posteriori* probability distribution $p_m = \frac{V_m}{\sum_{k=1}^{M} V_k}$ (with the summation taken over all patches with nontrivial occupancy). This is *Parker's* *matching principle* (Parker, 1978). For smaller discretized populations, expectations deviate to a limited degree from (6). When patch value is proportional to the *a priori* probability, $V_m = ca_m$, then the posterior and *a priori* probability distributions are again the same $p_m = a_m$ (up to discretization effects).

In the generic case, where the sequential process is decided without any ties, there is a deterministic allocation, and so the variance of the total on each patch is 0. With ties, some or all variances will be non-zero (e.g., see Figure 1).

Competitor avoidance follows a similar behavior. Here we assume that the utility functions are given by

$$U_m = \frac{V}{Y_m + 1} \tag{7}$$

where Y_m is the number of current occupants and V is the common value of all patches. In essence the population disperses uniformly over the locations (or as uniformly as discretization permits). If $N = kM$ for some integer k, then each patch will be occupied by exactly k individuals and the probability of a given individual going to a given patch is $1/M$. Moreover, if $kM < N < (k+1)N$ for some non-negative integer k, then all patches will have at least $k = \lfloor N/M \rfloor$ individuals and a patch P_m will have $k + 1$ individuals with probability $N/M - \lfloor N/M \rfloor$. Here $\lfloor \cdot \rfloor$ is the lowest integer value function. It follows that

$$E[X_m] = \frac{N}{M} \tag{8}$$

$$\mathrm{Var}(X_m) = \left(\frac{N}{M} - \left\lfloor \frac{N}{M} \right\rfloor\right)\left(\frac{M-N}{M} + \left\lfloor \frac{N}{M} \right\rfloor\right) \tag{9}$$

We note that here there are many ties in the sequential allocation process, so that the variances for all patches are non-zero, as we discussed regarding the ideal free distribution above.

2.1.4. *General utility function.* There are potentially many more plausible utility functions. For example, individuals may prefer to be in groups of reasonable size (for example wolves prefer to be in groups that are not to small so that they can catch prey, but not too large so that they do not have to share with too many individuals). Moreover, if we have $K \geq 1$ optimizing procedures with utility functions $U_m^{(k)}$ and K non-negative numbers $s_1, s_2, \dots s_K$ (potentially with $\sum_k s_k \neq 1$), we can easily combine them into a utility function

$$U_m = \sum_k s_k U_m^{(k)}. \tag{10}$$

The utility function considered in (10) reflects the trade-off costs in trying to achieve competing objectives such as social cohesion, maximization of individual space or resource usage, and site fidelity. The weighting terms s_k represent the relative strength of response for each goal's stimulus.

We should note that when combining utility functions, the specific functions need to be considered carefully. For example consider two patches, with possible utility functions $U_1^{(1)} = 3, U_2^{(1)} = 0$, $U_1^{(2)} = 1, U_2^{(2)} = 0$ and $U_1^{(3)} = 0, U_2^{(3)} = 2$. A population using $U^{(1)}$ will have all individuals go to patch P_1, as will a population using $U^{(2)}$. A population using $U^{(3)}$ will all go to patch P_2. An equal weighting of $U^{(1)}$ and $U^{(3)}$ will have all individuals going to patch P_1, whereas an equal weighting of $U^{(2)}$ and $U^{(3)}$ will have all going to patch P_2. Thus the final outcomes of the mixed cases are completely different, even though they used the same weighting of pure cases with identical outcomes.

2.2. **Probabilistic selection models.** In Section 2.1 we considered a range of models where individuals had a utility function and simply made the movement which maximized their utility at the time. Here we consider a stochastic version of these models, where individuals pick a patch with probability proportional to the utility function, i.e. a patch P_m is picked with probability $U_m / \sum_k U_k$.

We note that this mechanism is related to reinforcement learning (Sutton and Barto, 2018). In reinforcement learning, an individual has an assessment of the quality of different choices and updates this assessment through experience rather than simply picking what appears to be the best. As more data is accrued, an individual chooses the better options with higher probability.

As in Section 2.1, we distinguish several cases based on how the utility functions depend on the patch occupancy. We can visualize the probabilistic selection models in this section as urn models (Johnson and Kotz, 1977). Balls with numbers $1, 2, \dots, M$ are put in an urn and the balls are then sequentially drawn from the urn at random. The number of balls with the number m in the urn just before the n^{th} ball is drawn corresponds to the utility function U_m. If, the n^{th} ball drawn from the urn has a number m, the n^{th} individual moves into patch P_m. After the draw any of the following can happen:

1. The selected ball is returned to the urn (corresponding to the utility function independent of the occupancy).
2. The selected ball is returned to the urn and one extra ball with the same number is added to the urn (corresponding to the utility function positively correlating with the occupancy).
3. The selected ball is not returned to the urn (corresponding to the utility function negatively correlating with the occupancy).

We will see below that the above three are only the simplest examples and many more scenarios can be considered.

Finally, for any set of utility functions U_m, we can consider a family of utility functions

$$\widetilde{U}_m^{(s)} = \exp(sU_m) - 1 \tag{11}$$

parametrized by a sensitivity parameter $s \in (0, \infty)$ and place individuals proportional to $\widetilde{U}_m^{(s)}$. As $s \longrightarrow 0^+$, we have $\widetilde{U}_m^{(s)} \approx sU_m$ and thus our process would be the same as if we used functions U_m. As s grows, individuals prefer more and more the patch with maximal utility. In the limit of $s \longrightarrow \infty$, we thus recover the deterministic processes discussed in Section 2.1.

2.2.1. *Random selection—utility independent of the patch occupancy.* The most fundamental approach to patch selection is the assumption that each individual independently selects one location in accordance with a fixed probability distribution that applies to all members of the population. In this fully independent model (Broom and Rychtář, 2012), the probability that individual I_n moves to patch P_m is independent of the local population levels. Consequently, there is no distinction in outcomes whether settlement occurs simultaneously or sequentially.

We can achieve this distribution with utility functions

$$U_m = a_m \tag{12}$$

or by assuming a simple urn model with replacement with the urn having B balls in total and $a_m B$ balls with the number m.

The statistical analysis of identical independent random selection is straightforward: the expected number of individuals in patch P_m equals $E[X_m] = a_m N$ while the variance is $\mathrm{Var}(X_m) = a_m(1 - a_m)N$. Note that this means that $p_m = E[X_m]/N$, the posterior probability of a given individual being on patch P_m, is equal to a_m, i.e. this procedure is faithful. See Figure 1 for the illustration of the random process.

2.2.2. *The Polya urn—utility positively correlates with the patch occupancy.* In the classical Polya urn model, the selected ball is returned to the urn and one extra ball with the same number is added to the urn.

Here, we will consider a family of these models. For a parameter $B \in (0, \infty)$ corresponding to the initial number of the balls in the urn, we define utility functions

$$U_m = Ba_m + Y_m \tag{13}$$

where Y_m is the current occupancy at the patch P_m. Ba_m is the inherent site preference individuals have for location m, while the scaling parameter B moderates the population level at which density dependent social aggregation emerges. This is thus the stochastic model using the utility function similar to the follow the majority model.

If seen as an urn model, Ba_m corresponds to the initial count of balls with number m. It is known (Johnson and Kotz, 1977) that the distribution of the number of individuals at each patch follows the Dirichlet multinomial distribution with parameters $(N, Ba_1, \ldots, Ba_M)$. The expected value for X_m, the number

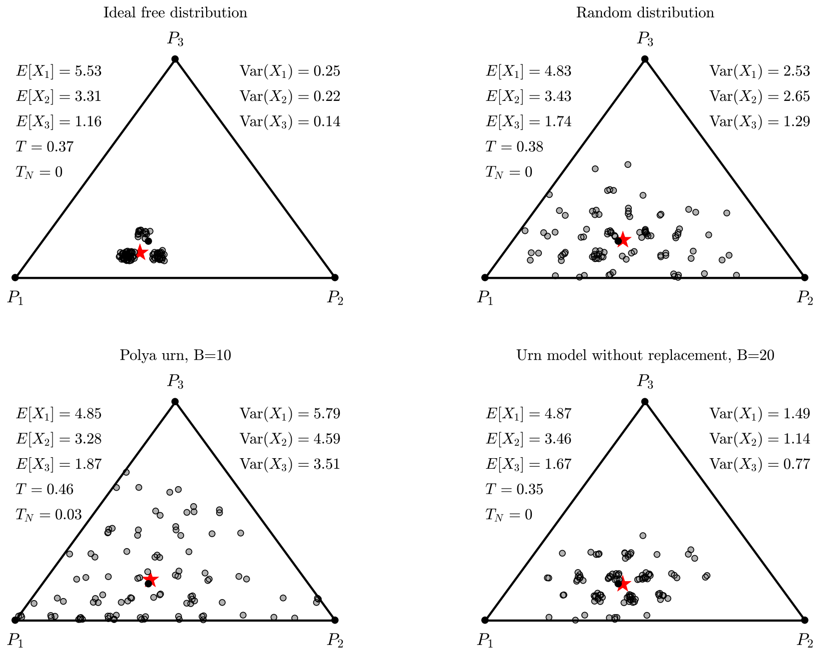

Figure 1. Example of 100 observations of movements of 10 individuals on 3 patches with $a_1 = 1/2$, $a_2 = 1/3$, $a_3 = 1/6$. The black disc is the predicted expected distribution, the red star is the simulated expected distribution. The procedure is faithful if the black disc and red star coincide. Every gray circle means one distribution from one simulation. Note that here we use a "shaky hand" method of drawing the points, so that rather than plotting the exact point, a small random error is added. This is because there are in fact a finite number of distinct possible values, and if the exact values were plotted it would not be possible to see the frequency of the occurrence of each. For every patch P_m we also provide the average number of occupants, $E(X_m)$, its variance, $\text{Var}(X_m)$, and movement coordination measures T and T_N (given by (25) and (27) as discussed later in Section 4) realized in the simulations.

of individuals in patch P_m, is given by

$$E[X_m] = N\frac{Ba_m}{\sum_k Ba_k} = Na_m. \tag{14}$$

Note that this means that $p_m = E[X_m]/N = a_m$, i.e. this procedure is faithful. Similarly, the variance of the number of individuals in a patch P_m is given by

$$\begin{aligned}\text{Var}(X_m) &= N\frac{Ba_m}{\sum_k Ba_k}\left(1 - \frac{Ba_m}{\sum_k Ba_k}\right)\frac{N + \sum_k Ba_k}{1 + \sum_k Ba_k} && (15)\\ &= Np_m(1 - p_m)\frac{N + B}{1 + B}. && (16)\end{aligned}$$

There is a nice closed form for the probability distribution as well (which we omit), which means that we can simulate the outcome of this distribution easily.

We note that although we have discussed this Section in terms of drawing balls from an urn, Ba_m does not have to be integer-valued for this process or the formulae to hold. For example, to keep B as a ball count, the expression $B \longrightarrow 0^+$ should be interpreted as adding a large number of balls (in the order of $1/B$) after each draw. This can be seen from the fact that utility functions $U_m = Ba_m + Y_m$ produce the same mechanism as $U'_m = a_m + Y_m/B$.

See Figure 1 for the illustration of this model.

2.2.3. *Drawing without replacement—utility negatively correlates with the patch occupancy.* In the classical urn model without replacement, the selected ball is discarded. Here, we will consider a family of these models. Consider a parameter $B \in (0, \infty)$ and define the utility function as

$$U_m = \max\{Ba_m - Y_m, 0\}. \tag{17}$$

If seen as an urn model, Ba_m corresponds to the initial count of balls with number m. We note again that although Ba_m represents a number of balls, it does not have to be integer-valued for this process or the formulae below to hold. We must ensure that U_m above does not become negative, however.

It is known (Johnson and Kotz, 1977) that the distribution of the number of individuals at each patch follows the multivariate hypergeometric distribution with parameters $(N, Ba_m, \dots, Ba_m)$. The expected value for X_m, the number of individuals in patch P_m, is given by

$$E[X_m] = N \frac{Ba_m}{\sum_k Ba_k} = Na_m. \tag{18}$$

Note that this means that $p_m = E[X_m]/N = a_m$, i.e. this procedure is faithful. Similarly, the variance of the number of individuals in a patch P_m is given by

$$\begin{aligned} \mathrm{Var}(X_m) &= N \frac{Ba_m}{\sum_k Ba_k} \left(1 - \frac{Ba_m}{\sum_k Ba_k}\right) \frac{\sum_k Ba_k - N}{\sum_k Ba_k - 1} && (19) \\ &= Np_m(1 - p_m)\frac{B - N}{B - 1}. && (20) \end{aligned}$$

See Figure 1 for the illustration of this model.

2.2.4. *General utility function.* As already briefly discussed in Section 2.2.2, the Polya urn model from can easily be modified by adding more than one ball with the replacement. The more balls we add, the more we “reinforce behavior” (Johnson and Kotz, 1977). We can also envision an urn model for which different patches reinforce differently. The Polya urn utility $U_m = Ba_m + Y_m$ can be adapted to $U_m = Ba_m + r_m Y_m$ for some $r_m > 0$; i.e. we generalize the Polya urn to the case when different patches reinforce their occupation at different rates. For example add 2 balls if patch 1 is selected and add just 1 ball for every other patch. How many balls are added may also depend not only on the "value" of the patch but also on how many individuals are already there, similarly to the optimizing models above, almost literally as animals (such as ants) mark their paths. There may be a threshold at which the balls are not added and perhaps not even replaced.

Similarly to Section 2.1.4, the utility functions do not have to be strictly decreasing or strictly increasing functions of the patch occupancy but can have (local) maxima and minima for intermediate occupancy levels.

Moreover, if we have $K \geq 1$ optimizing procedures with utility functions $U_m^{(k)}$ and K non-negative numbers $s_1, s_2, \ldots s_K$ (potentially with $\sum_k s_k \neq 1$), we can easily combine them into a utility function, as shown in equation (10). Note that combining utilities as in (10) is equivalent to considering the individual I_n going to patch P_m with probability proportional to $\sum_k s_k U_m^{(k)}$.

Finally, we can create a whole family of different models by appropriate combination of utility functions. For example, by setting

$$\begin{aligned} U_m(r,\theta) &= \exp\left(\frac{r\cos(\theta)}{1-r\cos(\theta)}(Y_m+1)\right) + \exp\left(\frac{-r\cos(\theta)}{1+r\cos(\theta)}\frac{1}{Y_m+1}\right)\ldots \\ &\ldots + \exp\left(\frac{r\sin(\theta)}{1-r\sin(\theta)}\frac{a_m}{Y_m+1}\right) + \exp\left(\frac{-r\sin(\theta)}{1+r\sin(\theta)}a_m\right) + a_m. \end{aligned} \tag{21}$$

with parameters r and θ, we can recover many of the models mentioned so far, see Figure 2 and Figure 3.

2.3. **Other sequential models.** In the previous sections, we assumed that individuals are trying to optimize the utility function either in some deterministic or probabilistic fashion. Here, we assume that individuals are selecting the patches sequentially and potentially in a random order, but individuals can observe the moves of their predecessors and then make their selection accordingly.

There are many possible scenarios, but we will consider only two. In the *follow the leader* model, the first individual will pick a patch based on the *a priori* probabilities a_m's and every other individual will pick the same patch.

In the *follow your predecessor* model, the first individual will pick a patch based on the *a priori* probabilities a_m's and every other individual will pick the same patch as its direct predecessor.

Note that under both of these models, all individuals will settle on patch P_m with probability a_m. This means both processes are examples of a herding process. Clearly, $E[X_m] = a_m N$, i.e. the procedures are faithful, and $\mathrm{Var}(X_m) = N^2 p_m(1 - p_m)$, i.e. the values are the same as in the follow the majority model

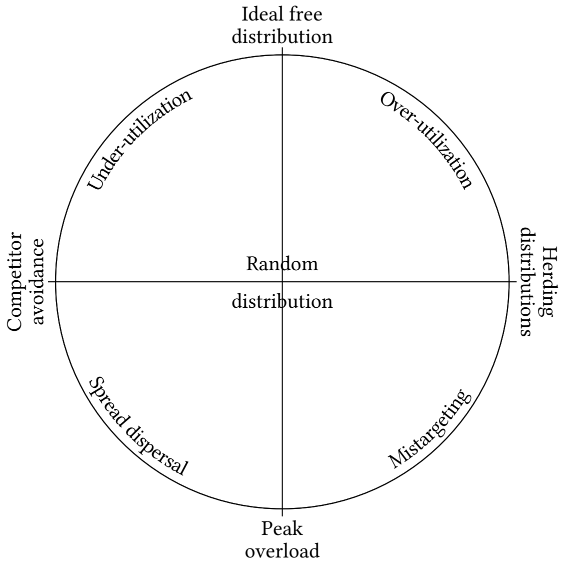

Figure 2. The utility function is parametrized by r and θ. Here, r represents the distance from the center, θ is the angle from the x-axis. For $r = 0$, we get the random distribution. For $r \longrightarrow 1^-$ we recover follow the majority (for $\theta = 0$), ideal free distribution (for $\theta = \pi/2$), competitor avoidance (for $\theta = \pi$), and peak overload (for $\theta = 3\pi/2$). See also Figure (3). Each point in the circle, (r, θ), represents a different model with a unique optimization goal (θ) and intensity in pursuit of that goal (r). See equation 21 for an example way of generating these (see also Section 3.2 below).

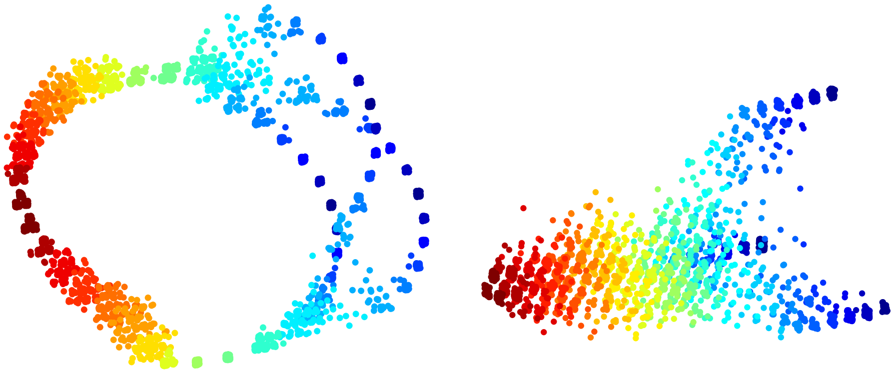

Figure 3. Illustration of combinations of different utility functions. Each panel represents 100 simulations of 10 individuals on 3 patches with $a_1 = 1/2, a_2 = 1/3, a_3 = 1/6$. (a) Tracing the outer circle of Figure 2 with utility functions given by equation (21) with $r = 0.99$ and varying θ. Starting at follow the majority (dark blue) for $\theta = 0$ and going counter-clockwise, we get ideal free distribution (green) for $\theta = \pi/2$, competitor avoidance (dark red) for $\theta = \pi$, and peak overload (green) for $\theta = 3\pi/2$. (b) Tracing the x-axis of Figure 2. competitor avoidance (dark red), random (green), follow the majority (dark blue) and everything in between.

from Section 2.1.2. We will see the differences between the three models in Section 3.2.2 when we will consider all these herding models in combination with other models; see also Figure 7 .

3. Novel models

The mechanisms considered in Sections 2.1–2.3 were generally known although perhaps not always applied to modelling of animal movement (such as the Polya Urn models). In contrast, the models presented here, in particular the "wheel" model described in Section 3.1, are novel models and new approaches to modeling animal movements. We also note that the wheel is a theoretical abstraction rather than a mechanistic model that simultaneously allocates individuals to patches in a way that satisfies desirable distributional properties.

3.1. **The wheel and base model.** Here, in contrast to the previous sections, we will not suppose any sequential underlying mechanism of animal patch selections, and instead we describe a procedure of simultaneous allocation of all individuals to patches. The methodology is easy to apply, and to visualize, and achieves the prescribed expected occupancy (i.e. is faithful) whilst allowing for varying degrees of randomness.

We visualize the procedure as follows, see Figure 4. A base disc of perimeter 1 is divided into M patches $P_1, \dots, P_M$ in the shape of wedges. The patch P_m corresponds to a wedge of circular length $p_m = a_m$ (i.e. a sector of angle $2\pi p_m$) so that $\sum_m p_m = 1$. The patches are ordered at random (see Figure 4(a)) to prevent any bias.

On top of the base disc is an upper disc, the wheel, comprising the N individuals represented by N spikes. The angle between individual I_i and I_j is $2\pi\theta_{ij}$ where $\theta_{ij} \in [-1/2, 1/2]$ is potentially drawn from a probability distribution (although of course not all angles between pairs can be independent; in fact once θ_{1j} are fixed for $1 < j \le N$, all other θ_{ij} are determined) and the signs correspond to a clockwise or anti-clockwise direction from I_i to I_j. Note that $\theta_{ij} = -\theta_{ji}$; see Figure 4(b).

Once the angles between the spikes are fixed, the wheel rotates by an angle φ chosen uniformly at random. The individual I_i then moves to patch P_m if and only if the corresponding spike finishes up on top of the corresponding patch; see Figure 4(c).

Using the wheel model, we can emulate outcomes of many of the procedures discussed above. For example, if, for all $1 < j \le N$, the angles θ_{1j} are chosen uniformly at random, we will recover random model from Section 2.2.1. By setting $\theta_{1j} = 0$, for all i, j, we recover the herding models from Section 2.1.2

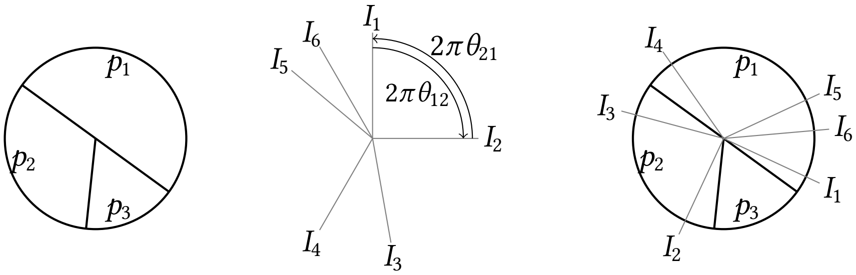

Figure 4. (a) Representing $M = 3$ patches with probabilities $p_1 = 1/2, p_2 = 1/3, p_3 = 1/6$. In general $\sum_m p_m = 1$. (b) Representing $N = 6$ individuals as spikes. The angle between individuals I_i and I_j is $2\pi\theta_{ij}$. In this case, $\theta_{12} = 1/4 = -\theta_{21}$. (c) After the upper disc with individuals is turned at a random angle over the base disc, the individuals land on patches. In this case, I_1, I_4, I_5 and I_6 will go to P_1, I_2 and I_3 will go to P_2, nobody will go to P_3.

and Section 2.3. By setting $\theta_{1j} = \frac{j-1}{N}$ we recover the competitor avoidance distribution from Section 2.1.3

By setting $\theta_{1j} \sim N(\mu_j, \sigma^2)$ to be a normally distributed random variable with mean μ_j and variance σ^2, we get a whole family of models. For $\mu_j = 0$ and small σ^2, we get models similar to herding, where almost all individuals are likely to end up at the same patch. As σ^2 grows, the tendency to go to the same patch is weakened as stochasticity has a significant effect, much as in the Polya urn. For large σ^2, the outcome resembles that of the random model; see Figure 5. Similarly, for $\mu_j = \frac{j-1}{N}$ and small σ^2, the model is close to the competitor avoidance distribution; as σ^2 grows, the increasing stochasticity leads to an outcome similar to drawing without replacement, with large σ^2 leading effectively again to the random distribution. See Figure 6.

3.2. **Hybrid models.**

We saw in Section 2.1.4 a way to combine several utility functions into one, and then in Section 2.2.4 how to combine several different procedures into a probabilistic one. In this section we investigate two more ways to combine the procedures discussed above.

3.2.1. *Hybrid type I.* Assume we have K procedures and K non-negative numbers $s_1, s_2, \ldots s_K$ with $\sum_k s_k = 1$. We can now consider a procedure defined as follows: with probability s_k, all individuals will follow the k^{th} procedure.

For example, different groups of individuals may use one of a repertoire of different movements, with the choice depending upon external factors not governed by the group (e.g., weather, the presence of predators).

This procedure allows for simple calculations. If $E[X_m^{(k)}]$ is the expected number of individuals in the patch P_m under the k^{th} procedure, and $\mathrm{Var}(X_m^{(k)})$ is the

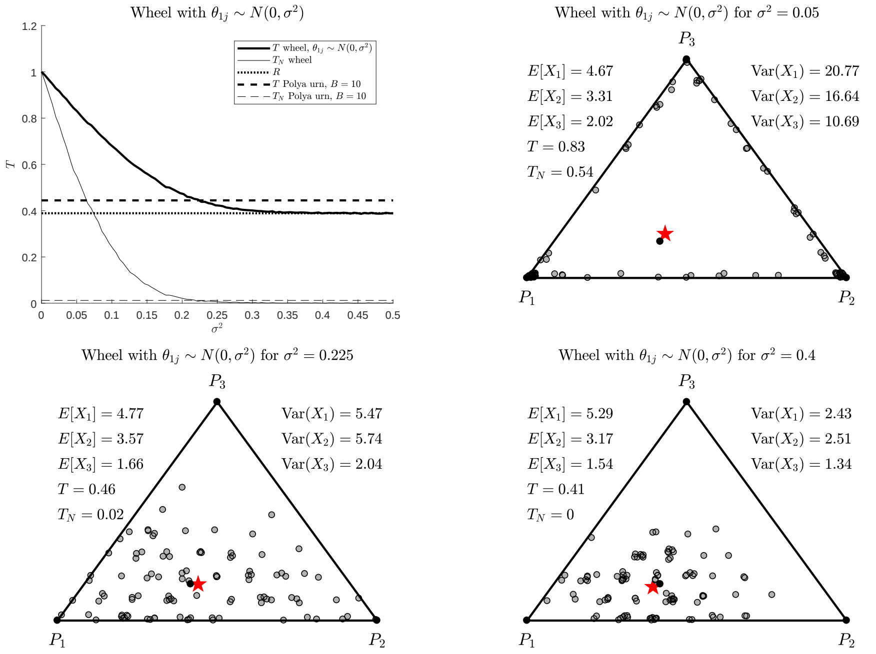

Figure 5. Simulation for the wheel with $\theta_{1j} \sim N(0,\sigma^2)$, $N = 10$, 100 observations, $a_1 = 1/2, a_2 = 1/3, a_3 = 1/6$. In the first figure, the comparison is done with the Polya urn model with $B = 10$. The other three figures show population distributions for the wheel with different values of σ^2.

variance of that quantity, we get

$$E[X_m] = \sum_k s_k E[X_m^{(k)}], \tag{22}$$

$$\mathrm{Var}(X_m) = \sum_k s_k \mathrm{Var}(X_m^{(k)}) + \sum_{j<k} s_j s_k (E[X_m^{(j)}] - E[X_m^{(k)}])^2 \tag{23}$$

In particular, when our original procedures are all faithful, the hybrid procedure will be faithful as well and (23) becomes $\mathrm{Var}(X_m) = \sum_k s_k \mathrm{Var}(X_m^{(k)})$.

3.2.2. *Hybrid type II.* Assume we have K sequential procedures and K non-negative numbers $s_1, s_2, \dots s_K$ with $\sum_k s_k = 1$. We can now consider a procedure defined as follows: each individual will follow the k^{th} procedure with probability s_k.

As seen from Figure 7, type I and type II hybrid methods lead to different combinations. Moreover, whilst the type I method is a probabilistic choice of which pure method to use, the type II method leads to a mixing of the individuals involved. This allows us to distinguish mechanisms that may look the same at first sight such as the three herding procedures. We also discuss this below in Section 3.3. Finally, we note that, in contrast to type I, a type II combination of faithful models does not have to be faithful, see Figure 7 showing the

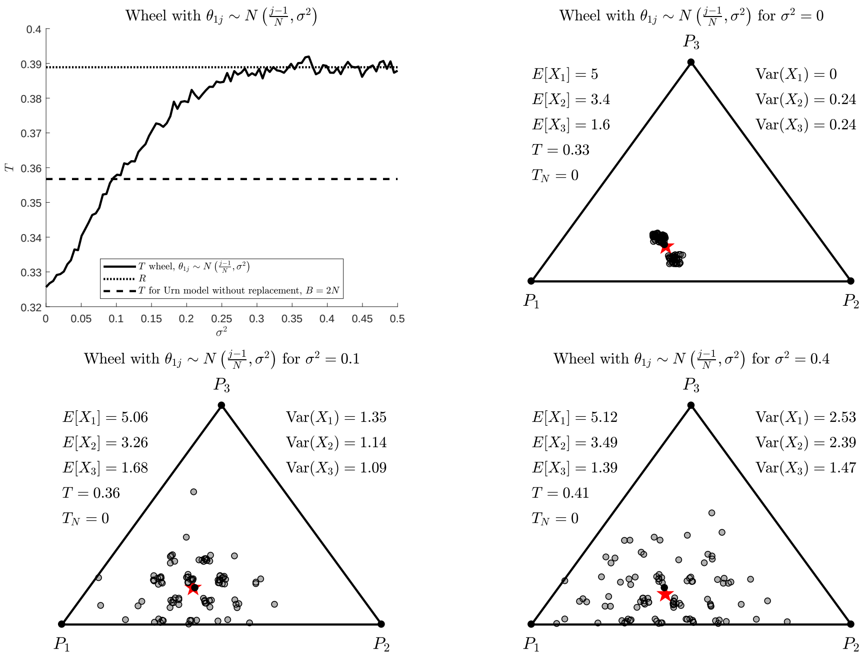

Figure 6. Simulation for the wheel with $\theta_{1j} \sim N\left(\frac{j-1}{N}, \sigma^2\right)$, $N = 10$, 100 observations, $a_1 = 1/2, a_2 = 1/3, a_3 = 1/6$. In the first figure, the comparison is done with the urn model with drawing without replacement for $B = 20$. The other three figures show population distributions for the wheel with different values of σ^2. The values of T_N are almost 0 for these parameters.

combination of random and follow the majority models. The fact that the hybrid type I and hybrid type II are different procedures is further demonstrated analytically in Section 4.4.2.

3.3. **Comparing models and outcomes.** We may be interested in determining if two models are effectively the same or not. We shall say that two models are *equivalent* if for any k-tuple of individuals and any patch, the probability of the k-tuple being on the given patch is the same under both models. In particular, as seen from the calculations in Section 4, two models are different if they produce different variances in the number of occupants on the patch.

Nevertheless, the above equivalence may not be strong enough as already seen for the example of herding models such as the follow the majority, follow the leader or follow the predecessor models. When these models are considered on their own, they yield completely identical outcomes. However, when considered in combination with a random movement under hybrid type II, they yield different outcomes; see Figure 7.

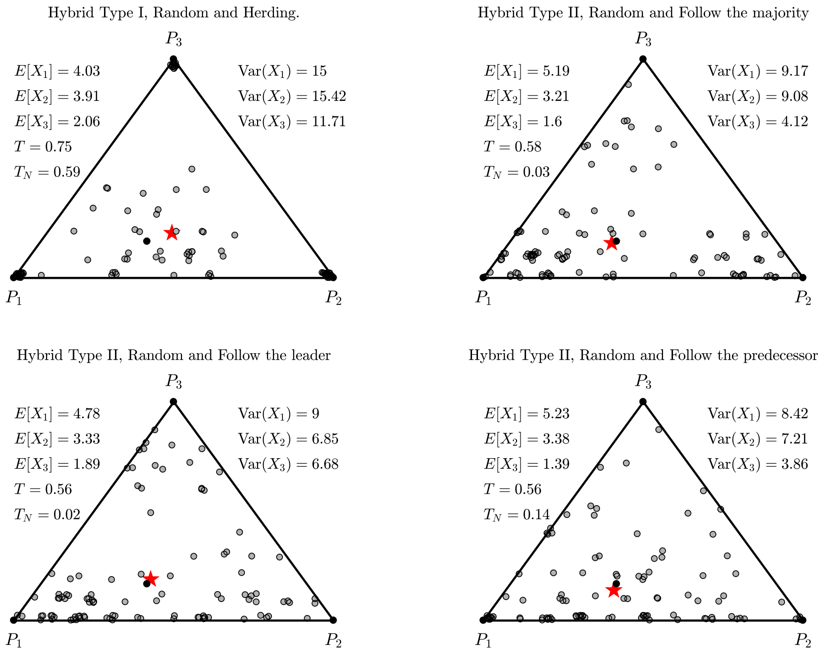

Figure 7. An example of 100 observations of movements of $N = 10$ individuals on 3 patches with $a_1 = 1/2$, $a_2 = 1/3$, $a_3 = 1/6$. The hybrid type I combinations, $s_1 = s_2 = 0.5$, of the random model with herding (follow the majority, follow the leader or follow the predecessor - give the same). Hybrid type II combinations, $s_1 = s_2 = 0.5$, of the same models differ. The combination with follow the majority will give bias towards patch P_1 (using random, most individuals will accumulate there and following the majority will only make the group bigger). The other two combination are faithful: the combination with follow the leader usually having most individuals at one or two patches, whilst that with follow the predecessor often has significant groups in all patches.

Note that we saw that in Section 2.1.4 a different way in which combining models that appear the same can lead to different results, when considering utility functions even in deterministic optimization models from Section 2.1.

3.3.1. *A fan and base model.* Here we show a movement that cannot be achieved by any of the above procedures. We can visualize the mechanism as in Section 3.1 but we will replace the wheel of spikes with fixed angles θ_{ij} by an expanding "fan" of spikes.

Consider two individuals on three patches with $a_1 = a_2 = a_3 = 1/3$. As in the wheel procedure, patches are represented as wedges on the disc (P_1 corresponding to a wedge from noon to 4 o'clock, P_2 from 4 to 8 and P_3 to a wedge from 8 to noon) and individuals as spikes. However, to describe a movement of the individuals, choose φ uniformly at random from $[0, 12)$ and move individual

I_1 to a corresponding hour while moving individual I_2 to an hour corresponding to 2φ. Then we have the following cases, each happening with probability 1/6:

- for $\varphi \in [0, 2)$, I_1 and I_2 will meet on patch P_1
- for $\varphi \in [2, 4)$, I_1 will be on P_1 and I_2 will be on P_2
- for $\varphi \in [4, 6)$, I_1 will be on P_2 and I_2 will be on P_3
- for $\varphi \in [6, 8)$, I_1 will be on P_2 and I_2 will be on P_1
- for $\varphi \in [8, 10)$, I_1 will be on P_3 and I_2 will be on P_2
- for $\varphi \in [10, 12)$, I_1 will be on P_3 and I_2 will be on P_3

So, individuals can meet on patch P_1 or P_3, but not on patch P_2. Such a situation is not possible under the wheel mechanism from Section 3.1. Indeed, under such a mechanism, if two individuals meet on patch P_m with probability μ_m and, if for some m' we have $a_{m'} \geq a_m$, then the same individuals have to meet on patch $P_{m'}$ with probability $\mu_{m'} \geq \mu_m$.

To show that this movement is not possible even for deterministic optimizing procedures from Section 2.1, realize that when there is an individual on patch P_1, the other is never on P_3 and vice versa. If there are some utility functions describing this kind of movement, we would need an empty patch P_2 being worth more than an empty patch P_1 or P_3 (so that when there is an individual on P_1, the other will end up on P_2 and not P_3). Consequently, when the first individual is to be placed, it will end up on patch P_2 and the second will end up on either P_1 or P_3. That is, they will never meet in P_1 nor P_3.

It is not possible to achieve this distribution under the probabilistic models from Section 2.2. There must be a nonzero probability for the first individual to move to patch P_1 or P_3 (otherwise, the individuals would never be able to meet at P_1 or P_3). Consequently, no matter how the utility depends on the occupancy, when the first individual moves to P_1, there is still a nonzero chance for the second individual to move to patch P_3. Consequently, there is a nonzero chance that the individuals will end up in patches P_1 and P_3 at the same time.

We can still describe this mechanism in terms of deterministic or probabilistic optimizing of the utility functions similarly to what is done in Sections 2.1 and 2.2. Nevertheless, we will have to allow for the utility functions to depend on the occupancy of other patches (e.g., when a predator of our population can observe all patches and choose which to attack depending upon occupancy). One possible mechanism is as follows. If all patches are empty, the utility of each patch is 1/3. If the patch P_1 is occupied, the utilities are given by $U_1 = 1/2, U_2 = 1/2, U_3 = 0$. If the patch P_2 is occupied, the utilities are given by $U_1 = 1/2, U_2 = 0, U_3 = 1/2$. Finally, if patch P_3 is occupied, utilities are given by $U_1 = 0, U_2 = 1/2, U_3 = 1/2$. With the utility functions defined this way, we can then follow the methods in Section 2.1 or Section 2.2 to achieve the desired movement.

4. NOVEL MEASURES OF MOVEMENT COORDINATION

We consider a group of individuals making a single movement following a process described above. We shall define the following novel measures of movement coordination. Let T_{ij} denote the probability that individuals I_i and I_j are on the same patch as each other. The probability of a randomly selected pair of individuals being on the same patch is then

$$\begin{aligned} T_2 &= \frac{1}{\binom{N}{2}} \sum_{i<j}^{N} T_{ij} && (24) \\ &= \frac{1}{N(N-1)} \sum_{m=1}^{M} E[X_m(X_m-1)]. && (25) \end{aligned}$$

where X_m denotes the number of individuals on patch P_m.

Similarly, let $T_{i_1\ldots i_k}$ denote the probability that the k individuals $I_{i_1}, \ldots, I_{i_k}$ are all on the same patch. The probability of a randomly selected group of k individuals being on the same patch is then

$$\begin{aligned} T_k &= \frac{1}{\binom{N}{k}} \sum_{i_1<i_2<\ldots<i_k}^{N} T_{i_1\ldots i_k} && (26) \\ &= \frac{1}{\binom{N}{k}} \frac{1}{k!} \sum_{m=1}^{M} E[X_m(X_m-1)\cdots(X_m-k+1)]. && (27) \end{aligned}$$

In the following, we will simply denote T_2 as T.

Using $E[X_m] = p_m N$, we get

$$\begin{aligned} T &= \frac{1}{N(N-1)} \sum_m E[X_m(X_m-1)] && (28) \\ &= \frac{1}{N(N-1)} \sum_m \Big(\mathrm{Var}(X_m)) + \big(E[X_m]\big)^2 - E[X_m]\Big) && (29) \\ &= \frac{1}{N(N-1)} \sum_m \Big(\big(Np_m\big)^2 - Np_m\Big) + \frac{1}{N(N-1)} \sum_m \mathrm{Var}(X_m) && (30) \\ &= \frac{\big(N\sum_m p_m^2\big) - 1}{N-1} + \frac{1}{N(N-1)} \sum_m \mathrm{Var}(X_m) && (31) \end{aligned}$$

Consequently, T is simply a function of N, $p_m = E[X_m]/N$, and $\mathrm{Var}(X_m)$. These are perhaps the most fundamental properties of any movement process. However, it is useful to have a single measure of aggregation, and T is the most natural one if the relationship between the individuals is the focus. We saw in

Section 2 a number of distinct mechanisms. The values of T and T_N for these mechanisms are shown in Table 1.

4.1. **Random movement—no coordination.**

We now evaluate T_k for the random model of Section 2.2.1. We will use this, and especially the result for T_2, for comparison with other models. For random movement, where individuals move to patch P_m with probability p_m independently of each other, we will denote the probability of a randomly selected group of k individuals being on the same patch by R_k, and we denote R_2 simply by R. The individuals meet at a patch P_m with probability p_m^k (note that there can be other individuals there too) and so

$$R_k = \sum_{m=1}^{M} p_m^k. \tag{32}$$

Thus we have $R = \sum_{m=1}^{M} p_m^2$. Alternatively, we can derive the above from (31) and the fact that $\mathrm{Var}(X_m) = Np_m(1 - p_m)$.

4.2. **Bounds on T.**

It is clear that for any k, $T_k \leq 1$. To find the lower bound on T, note that it follows from (31) that

$$T = \frac{NR - 1}{N - 1} + \frac{1}{N(N-1)} \sum_m \mathrm{Var}(X_m). \tag{33}$$

To minimize T is thus the same as to minimize $\sum_m \mathrm{Var}(X_m)$.

Assume that $Np_m = E[X_m]$ is not an integer for all m. Denote $f_m = Np_m - \lfloor Np_m \rfloor$ (and so $\lceil Np_m \rceil - Np_m = 1 - f_m$).

Let $(\widetilde{X}_m)_{m=1,\ldots,M}$ denote the multivariate distribution that minimizes $\sum_m \mathrm{Var}(X_m)$ while satisfying $E[\widetilde{X}_m] = p_m N$. We show in Appendix A that all of the M univariate distributions $\widetilde{X}_m$ must take only values Np_m (if this is an integer), or only values $\lfloor Np_m \rfloor$ and $\lceil Np_m \rceil$ (if Np_m is not an integer). Moreover, in the latter case, the probability $\widetilde{X}_m$ being $\lfloor Np_m \rfloor$ is given by $\frac{\lceil Np_m \rceil - Np_m}{\lceil Np_m \rceil - \lfloor Np_m \rfloor}$ which equals $\lceil Np_m \rceil - Np_m = 1 - f_m$, because if Np_m is not an integer, $\lceil Np_m \rceil = \lfloor Np_m \rfloor + 1$. The above distributions will be referred to as the minimal range distribution (for the given mean Np_m).

When the $\widetilde{X}_m$ are as above, we get

$$\begin{aligned}
E[\widetilde{X}_m(\widetilde{X}_m - 1)] &= \Big((1 - f_m)\lfloor Np_m \rfloor(\lfloor Np_m \rfloor - 1) + f_m \lceil Np_m \rceil(\lceil Np_m \rceil - 1)\Big) && (34)\\
&= 2Np_m \lfloor Np_m \rfloor - \lfloor Np_m \rfloor(\lfloor Np_m \rfloor + 1)) && (35)\\
&= \lfloor Np_m \rfloor(Np_m - \lceil Np_m \rceil) && (36)\\
&= N^2 p_m^2 - Np_m + f_m(1 - f_m) && (37)
\end{aligned}$$

and consequently

$$T = \frac{\sum_m E[X_m(X_m - 1)]}{N(N-1)} \geq \frac{\sum_m E[\widetilde{X}_m(\widetilde{X}_m - 1)]}{N(N-1)} \tag{38}$$

$$= \frac{\sum_m N^2 p_m^2 - N p_m + f_m(1 - f_m)}{N(N-1)} = \frac{RN^2 - N + \sum_m f_m(1 - f_m)}{N(N-1)} \tag{39}$$

We will denote the minimal value $\frac{RN^2 - N + \sum_m f_m(1-f_m)}{N(N-1)}$ of T by $T_{\min}$.

4.3. **Maximal and minimal values of T.** The certainty for a random pair of individuals to meet can be achieved if (and only if) all individuals move together in one big group (such as in the herding model), i.e. when there is maximal coordination between the individuals.

The minimal value of T, $T_{\min}$, can be achieved if individuals want to avoid each other. This can be modeled as a wheel with uniformly distributed spikes, the *regular wheel*. That we can achieve this distribution is demonstrated in Appendix B. We note that the regular wheel distribution is not the only one that gives this lower bound, as we also show in Appendix B.

We note that both large and small T values can be considered as a sign of a high degree of coordination, either high aggregation or high dispersal, respectively. Minimal coordination is associated with random movement and $T = R$, although there are cases with $T = R$ where there is coordination. This is similar to the connection between independence and correlation. Independence implies no correlation, but not vice versa.

4.4. **Intermediate values of T.** Using the hybrid type I model from Section 3.2.1, we can easily see that any value of T between $T_{\min}$ and 1 can be achieved. Indeed, pick $s \in [0, 1]$ and let all individuals move like a herd with probability s and use the regular wheel with probability $1 - s$. Under such a model, we will get $T = s + (1 - s)T_{\min}$, i.e. we can choose the parameter s so that T attains any value in $[T_{\min}, 1]$.

It is interesting that any value of T can be achieved through such a simple averaging mechanism of two of the most extremely coordinated systems. This should act as a reminder that T will not tell us everything, and to fully understand the coordination within a population, the higher moments T_k are needed also.

4.4.1. *Intermediate values of T for Polya urn models.* Consider the Polya urn model from Section 2.2.2 with parameter B. We get, by (14) and (16),

$$\begin{aligned} T &= \frac{1}{N(N-1)} \sum_m E[X_m(X_m-1)] && (40)\\ &= \frac{1}{N(N-1)} \sum_m \left(\mathrm{Var}(X_m) + (E[X_m])^2 - E[X_m]\right) && (41)\\ &= \frac{1}{N(N-1)} \sum_m \left(Np_m(1-p_m)\frac{N+a}{1+a} + N^2p_m^2 - Np_m\right) && (42)\\ &= \frac{1}{N(N-1)} \left(N(1-R)\frac{N+B}{1+B} + N^2R - N\right) && (43)\\ &= R + \frac{1-R}{1+B}. && (44) \end{aligned}$$

Thus T depends only upon B and the equivalent probability for the corresponding independent model, R. It ranges down from 1 for $B \longrightarrow 0^+$ to independent R with $B \longrightarrow \infty$ (technically with this method we cannot attain either bound).

We see that we can achieve any value of T between the purely random value R and the value that would be obtained through the deterministic herding procedure of Section 2.1.2 or Section 2.3. Using a similar approach, this will also be true for the probabilistic versions of any other optimization process based upon a utility function.

4.4.2. *Intermediate values of T: comparing hybrid models.* In this section we will show another example of how intermediate values of T can be attained and also further investigate the differences between hybrid model type I and hybrid model type II from Section 3.2.

First, as in hybrid type I, assume that with probability s_1^I all individuals will use the follow the leader model and, with probability $s_2^I = (1 - s_1^I)$, move at random (i.e., we apply the hybrid type I model). Thus, we get, for $k = 2, \dots, N$,

$$T_k = s_1^I + (1 - s_1^I) \sum_m p_m^k \qquad (45)$$

which ranges from $R_k = \sum_m p_m^k$ for $s_1^I = 0$ to 1 for $s_1^I = 1$.

Second, as in hybrid type II, assume that each individual uses the follow the leader model with probability s_1^{II}, and with probability $s_2^{II} = 1 - s_1^{II}$ the individual uses random (i.e., we apply the hybrid type II model).

For any randomly selected pair of individuals, there is a probability of $2/N$ that one of them is the leader, in which case the probability they are together is $s_1^{II} + (1 - s_1^{II}) \sum_m p_m^2$. Otherwise, with probability $(N-2)/N$, the following three cases can happen:

1. With probability $(s_1^{II})^2$ both individuals followed the leader thus ending up in the same patch for sure.
2. Exactly one individual followed the leader and the other was placed at random. This happens with probability $2s_1^{II}(1 - s_1^{II})$.
3. No individual followed the leader. This happens with probability $(1 - s_1^{II})^2$.

In cases 2 and 3, the individuals move independently of each other and so their probability to end up in the same patch is $\sum_m p_m^2$. Either case 2 or case 3 happens with probability $1 - (s_1^{II})^2$. Consequently, the probability for both individuals to end up in the same patch is given by $(s_1^{II})^2 + (1 - (s_1^{II})^2) \sum_m p_m^2$. Combining the two, we thus obtain

$$T = \frac{2}{N} \left(s_1^{II} + (1 - s_1^{II}) \sum_m p_m^2 \right) + \frac{N-2}{N} \left((s_1^{II})^2 + (1 - (s_1^{II})^2) \sum_m p_m^2 \right) \tag{46}$$

$$= \left((s_1^{II})^2 - \frac{2s_1^{II}(1 - s_1^{II})}{N} \right) + \left(1 - (s_1^{II})^2 + \frac{2s_1^{II}(1 - s_1^{II})}{N} \right) \sum_m p_m^2, \tag{47}$$

which again ranges from $R = \sum_m p_m^2$ for $s_1^{II} = 0$ to 1 for $s_1^{II} = 1$.

To evaluate T_N, note that N individuals can meet on patch P_m when the leader goes to patch P_m, k individuals follow the leader and the remaining $N - k - 1$ individuals end up on the patch P_m "by chance". This gives

$$T_N = \sum_m \binom{N-1}{k} (s_1^{II})^k (1 - s_1^{II})^{N-k-1} p_m p_m^{N-k-1} \tag{48}$$

$$= \sum_m p_m (s_1^{II} + (1 - s_1^{II}) p_m)^{N-1}. \tag{49}$$

The difference between type I and type II is that in type I, with probability larger than s_1^I, all individuals will follow the leader, while in type II a random number of individuals will follow the leader (the number of such individuals will follow a binomial distribution with parameter s_1^{II}). Thus, in type I, there is a group of size N with probability larger than s_1^I. However, in type II, the group size will be larger than for independent movement, but in a different, less extreme, way than type I.

For a fair comparison, choose s_1^I for type I and s_1^{II} for type II to satisfy (for large N) $s_1^I \approx (s_1^{II})^2$. Letting N tend to infinity, type I would yield $T_N = s_1^I$ and type II would yield $T_N = 0$ whilst T is the same for both models.

See Figure 7 for visual illustration of the differences between the two types of combinations.

4.5. **Model statistics.** We consider model statistics for each of the models that we have described in this paper in Table 1. We observe a wide variety of values of our statistics, depending upon the underlying mechanisms.

We see from the table that there are various ways that groupings of the entire population ($T_N = 1$) can result, and just observing such a group does not reveal the mechanism, which might relate to preference for grouping with others or just for the site itself. This can only be distinguished by repeated observation to see if different sites are selected. Conversely, knowledge of the variability of the group size at any given site is not sufficient to understand whether the site or the presence of others is the key factor in to what extent individuals aggregate.

The entries in Table 1 are for the pure models only, and a great variety of different results can be obtained through the hybrid models of Section 3.2. We observed in Section 4.4.2 that the pairwise association measure T_2 can be the same for different distributions, but that this may lead to large differences in the probability of the population forming a single group. We further note that mechanisms which appear identical in their outcome when followed with absolute accuracy, can in fact be very different when some error is introduced as we see in the three herding variants in Figure 7. Thus great care is needed when making inferences about mechanisms from the underlying data.

Mechanism	$p_m = \frac{E[X_m]}{N}$	$\mathrm{Var}(X_m)$	T_2	T_N
Peak overload	0 if V_m not max 1^* otherwise	0^*	1	1
Herding (Follow the majority/leader/predecessor)	a_m	$N^2 p_m(1 - p_m)$	1	1
Ideal free distribution	$\frac{V_m}{\sum_k V_k}$	0^*	$\frac{N \sum V_m^2 - (\sum V_m)^2}{(\sum V_m)^2 (N-1)}{}^*$	0^*
Competitor avoidance	$\frac{1}{M}$	$f_M(1 - f_M)$	$\frac{N - M}{M(N-1)} + \frac{M}{N(N-1)} f_M(1 - f_M)$	0
Random	a_m	$p_m(1 - p_m)N$	$R = \sum_m p_m^2$	$R_N = \sum_m p_m^N$
Polya urn with B balls	a_m	$N p_m(1 - p_m)\frac{N + B}{1 + B}$	$R + \frac{1 - R}{1 + B}$	$\sum_m \prod_{n=0}^{N-1} \frac{p_m B + n}{B + n}$
Drawing without replacement	a_m	$N p_m(1 - p_m)\frac{B - N}{B - 1}$	$R - \frac{1 - R}{B - 1}$	$\sum_m \prod_{n=0}^{N-1} \frac{p_m B - n}{B - n}$
Wheel	p_m	$0 \longrightarrow 1$	$T_{\min} \longrightarrow 1$	$0^* \longrightarrow 1$

Table 1. Summary statistics of different mechanisms. Several mechanisms can yield the same value for a given measure and a combination of measures is required to identify any given mechanism. $f_M = \frac{N}{M} - \left\lfloor \frac{N}{M} \right\rfloor$. *except for particular or extreme cases.

5. Discussion

The distribution of animal populations — whether in their selection of sites for colonization or as transient elections during migrations — has been a central question of ecology and its supporting mathematical literature. Here we have explored both existing and novel mechanisms governing patch selection and analyzed the degree to which each interpretation maintained the faithfulness between *a priori* and *a posteriori* probabilities. Although many mechanisms lead to the same or similar probabilities as statistical expectations, they can exhibit strikingly different realizations in practice.

The model that we have developed captures this feature. We can see this from Figure 7, where three mechanisms that are identical in outcome from the statistical measures in Table 1 are very different when there is some additional random effect, perhaps caused by errors or unpredictable external influence. The modelling methodology that we have developed can model a wide range of movement mechanisms and also produce mechanisms which have desirable features. For instance we may be interested in modelling a population with a certain level of aggregation, and we can specify values of our population measures and then find a mechanism that yields these (although as we see in Section 3.3.1 it is not clear that existing models can generate every possible case). This will enable us to use these methods as components of models with a wider range of features, such as from the evolutionary framework of Broom and Rychtář (2012). The primary focus of this paper, however, is the modelling of movement of populations itself, and we discuss this in detail below.

Long distance migration is an important aspect of the life cycle for many organisms including not only avian species (Hutto, 2000; Sillett and Holmes, 2002; Mehlman et al., 2005; Moore et al., 2005), but also fish (Ward et al., 2001), ungulates (Guttal and Couzin, 2010), primates (Pyritz et al., 2011), and whales (Heimlich-Boran, 1988). Both assessment of the landscape and individual memories from previous migration cycles affect the stopover locations for these species, with the latter providing a justification for both the existence of *a priori* probabilities and the faithfulness of *a posteriori* ones. For example, migratory ducks are believed to assess local conditions and adjust layover times in response to resource availability and their exposure to predation risk (O'Neal et al., 2012), while pods of killer whales make use of subsurface topography to navigate back to historic hunting grounds, although the main feeding grounds are themselves separated by deep featureless zones (Heimlich-Boran, 1988). Philopatry is also evident in female-led groups of red-faced lemurs (Pyritz et al., 2011).

The spatial co-occurence of animals in either migratory stopovers or during colonization and settlement of territories invites the question as to whether this spatial association was achieved via some mechanism of social cohesion or density dependent selection (congregation/schooling), or whether it was solely a product of density independent processes (aggregration/shoaling). The large herds of lion prey species on the African savannah act as a defense mechanism against predation (Hayward and Kerley, 2005; Scheel, 1993; Cressman and Garay, 2011), illustrating the positive herd reinforcement mechanism in our framework models. In contrast, stem mothers of the gall-forming aphid *Pemphigus betae* almost exclusively settle the largest available leaves irrespective of the fitness costs imposed by high competitor densities (Whitham, 1978), mirroring our concept of peak overload.

We believe Markov models will have a role in the study of aggregation and dispersal. It may be useful to conduct a field experiment for migratory birds, or the aphids *Pemphigus betae*, to estimate the utility functions and subsequently compare the predictions of our theoretical models to the outcomes of biological experiments.

Leaders are often individuals who are more attuned to environmental information and are less influenced by the aggregation of conspecifics (Guttal and Couzin, 2010). They make up a comparatively small proportion of groups relative to socially-focused individuals whose movements are socially facilitated (King, 2010; Petit and Bon, 2010; Pyritz et al., 2011). In our models, patch settlement was adopted sequentially but with the assumption that all members could assume the role of early adopters or leaders. This type of egalitarianism is evidenced in red-fronted lemurs, *Eulemur rufifrons*, which dwell in the Kirindy Forest of Madagascar (Pyritz et al., 2011). In more socially despotic lemurs and other primates with strong dominance hierarchies (Watts, 1994; Chapman, 1990), leadership is more stable and responsive to external threats such as the territorial encroachment of rival bands or out-group mate raiding opportunities. In our models, this would represent the more dominant individuals non-randomly appearing first in the sequencing for patch selections.

High levels of density-dependent movement are often associated with large variations in patch qualities (Gadgil, 1971; McPeek and Holt, 1992; Travis et al., 1999; Metz and Gyllenberg, 2001; Poethke and Hovestadt, 2002; Matthysen, 2005; Hixon et al., 2002; Ford and Swearer, 2013). The most common scenario is when settlement is negatively dependent (alternatively, dispersal is positively dependent) as a consequence of competition for limited resources with colonization rates proportional to the variation in patch quality (Fretwell and Lucas, 1969; Morris, 2003; Resetarits Jr and Binckley, 2009). Thus when beavers, *Castor canadensis*, were returned to historic, pre-trapping areas, they demonstrated clear preferences for areas in the lower watershed (Cunningham et al., 2006). Sites higher in the watersheds were settled last and abandoned first during field observations. Likewise, during stopover periods, migratory passerines have extreme energy demands as they attempt to replenish expended fat stores and exhibit territoriality, although conspecifics can aid in the identification of good sites (Moore and Yong, 1991). Nevertheless, movement and colonization can also respond positively to population density (our herding model) if there is an increased efficiency in foraging (Takahashi et al., 2004) or anti-predator benefits. In a secondary effect, dispersers from these larger aggregations may have an elevated risk of attack by peripheral predators attracted by the group (Matthysen, 2005). There may be other, social contexts that also reinforce the spatial association of in-groups versus out-groups, i.e., social fence theory (Hestbeck, 1982). For the southern hulafish, *Trachinops caudimaculatus*, a shoaling zooplanktivore off the southeastern coast of Australia, both resource/refuge competition and social cohesion are evident (Ford and Swearer, 2013). The hulafish responds to any patchiness within the reef structure by distributing in accordance with reef size; however, on any given reef the fish comprise a single shoal. This leads to possibly suboptimal distributions on large continuous reef structures.

The foraging behavior of Adélie penguins offers an interesting case study for the modeling framework presented here as there appears to be two processes at play. Adélie penguins, *Pygoscelis adeliae*, are known to dive synchronously into small open waters surrounded by sea ice (Wilson et al., 1986; Tremblay and Cherel, 1999; Ainley, 1972, 2002; Takahashi et al., 2004), possibly as a means to reduce predation risk upon entry into the water (Todd, 1988; Ainley, 1972, 2002; Rogers and Bryden, 1995). Diving is typically proceeded by a congregation of penguins on the ice followed by calls that are hypothesized to facilitate the dive synchronization (Takahashi et al., 2004). Diving in pairs or trios during the observations, the action is usually initiated by the same lead individuals (Takahashi et al., 2004). Once in the water, however, the penguins forage independently and disperse to different depths in the water column. This dispersal reflects the fact that the penguins' prey density is relatively small but occurs over a greater range of depth (O'Brien, 1987). Compare this behavior with rockhopper penguins who do not forage independently and whose prey occurs in dense swarms (Tremblay and Cherel, 1999). Consequently, the timing of diving seems similar to the wheel model with a very tight range of dive times, while the depth of foraging is controlled by a quality function negatively affected by the presence of conspecifics.

One of the framework features we examined was the transitioning between alternative utility objectives (e.g. Figures 2 and 3). The desert locust *Schistocerca gregaria* exhibits density-dependent phenotypic plasticity that affects both its social behavior (solitary crypsis versus aposematic gregarious swarming) and food source selection (avoidance or consumption of the toxic plant alkaloid hyoscyamine) (Uvarov, 1977; Simpson et al., 2001; Sword and Simpson, 2000; Sword and Dopman, 1999; Despland and Simpson, 2005). At low populations, solitary *S. gregaria* are difficult to locate for predators; however, when densities increase, the physical contact between individuals stimulates gregarization (Collett et al., 1998; Simpson et al., 2001). The resulting swarms are easily identified because the accompanying switch in diet leads to a bright colorization in addition to their numeracy; however, these locusts also acquire a measure of anti-predator protection. Predators learn to avoid the swarming locusts because they are rendered unpalatable following the consumption of the toxic compounds. Although the population initiates the transition from being solitary to gregarious in response to higher densities, the window of transition does not itself appear to be density-dependent.

Not all movements need pertain to consumable resources. When white storks, *Ciconia ciconia* migrate from Germany to Spain, the presence of conspecifics in flight can provide information on the location and strength of updrafts (Van Loon et al., 2011; Pennycuick, 1972). Some of these migratory populations can be partitioned into identifiable subgroups (Mueller et al., 2013; Scheiber et al., 2013; Voelkl and Fritz, 2017), but discernable social relations are not a requirement for large aggregation (Leshem and Yom-Tov, 1998), not unlike traveling wildebeest herds. Even when subgroups are present, however, their memberships are not static but change dynamically over the course of the long migration (cf. migrations of the red-faced lemurs described above).

Population movement and colonization have been addressed in a number of studies that have focused on particular mechanisms governing movement processes. The original ideal free distribution (Fretwell and Lucas, 1969) has led to modern interpretations of the concept using explicit spatial structures both discrete (Armsworth and Roughgarden, 2005) and continuous (Cosner, 2005; Rowell, 2009). The resulting distributions are qualitatively recapitulated in our model when the utility functions are negatively density-dependent. The continuous system of Reding et al. (2016) provides a point of contrast with our herding model. In both cases there is a general increase in the attractiveness of locations with rising density. With the herding model, the increase is unbounded, but in Reding et al. (2016) ecological demographic processes eventually bound this increase and limit density. Intermediate population sizes are also featured in more recent trends in research involving diffusive movement that uses U-shaped dispersal rules (e.g. Kim et al., 2009) where there is high dispersal when the population density is either low or high but not when it is intermediate. Travis et al. (1999) used a model of dispersal on $n \times n$ lattice networks to show that density dependent dispersal should evolve under most conditions, while Ballerini et al. (2008) demonstrated that collective animal behavior could emerge from simple rules of interactions between individuals (see also Guttal and Couzin, 2010). Moreover, what seemed to matter most was the topological (social) distance between individuals rather than the metric (physical) distance. Our present framework does not contain this level of social networking between individuals, but it could be extended to do so. Given the community structure and hunting strategies of cooperative predators like killer whales who sometimes use line-abreast positioning like an expanded net (Heimlich-Boran, 1988), reminiscent of our wheel model, this is a logical next step for our paper's framework.

In general it would be interesting to know to what extent the wheel can describe the different cases that optimizing and reinforcement models can do. We saw in Section 3.1 with Figures 5 and 6 that it is flexible it showing a lot of these behaviours, yet at the same time is a relatively simple model to work with. Perhaps we can generate a lot of what we want with just a single mechanism? We note also similar flexibility from the hybrid methods of Section 3.2.

There are many reasons why individuals elect to be part of a group (spatially, behaviorially) or to remain solitary (Krause and Ruxton, 2002), and we have demonstrated here a broad framework of sequential behavioral adaption that captures both alternatives. Many other factors influence these behavioral expressions, however, and future work should incorporate additional social and ecological realism in the formulation of our utility functions or hybrid model creation. We have previously mentioned the inclusion of social structure as one research direction. As a further consideration, many species exhibit sex-specific or age-specific segregation (Baguette et al., 1998; Matthysen, 2005; Dingle and Drake, 2007; Belthoff and Gauthreaux Jr, 1991), while other species may be influenced by the prevalence of exogenous and endogenous mating. We can simulate these effects using our framework by adding the appropriate dependence to the utility functions and tracking sub-group membership. This will necessitate further consideration of the role of sequencing the population in our model as sub-group interactions may have confounding effects. Additionally we have assumed that one patch does not influence another, yet perceived habitat quality can be influenced both positively and negatively by neighboring locations (Resetarits et al., 2005) (situations without this assumption were discussed in Broom and Rychtář, 2018) as mentioned in Section 2.1.

Finally, predator behavior strongly influences the aggregation of a population (Cressman and Garay, 2011), and the predators themselves can disperse across the environment. Our model offers the possibility of studying the dynamics of settlement between prey and predator species on both ecological and evolutionary timescales.

Appendix A. Proof of the minimal value of T

Suppose that the distribution of individuals at patch P_m, denoted by Q_m, is given by

$$P[Q_m = k] = q_{mk}, k \in \mathbb{N}_0. \tag{50}$$

We know that

$$T = \frac{1}{N(N-1)} \sum_{m=1}^{M} \Big(\mathrm{Var}(Q_m) + \big(E[Q_m]\big)^2 - E[Q_m] \Big) \tag{51}$$

$$= \frac{NR-1}{N-1} + \frac{1}{N(N-1)} \sum_{m=1}^{M} \mathrm{Var}(Q_m), \tag{52}$$

where

$$\mathrm{Var}(Q_m) = \sum_k q_{mk}(k - Np_m)^2. \tag{53}$$

We wish to find the minimum possible summation of these variance terms. We proceed to find the minimum for each individually, and show that these are achieved by the distribution generated by the minimal range distribution.

If Np_m is an integer then clearly the distribution with mean Np_m yielding the minimum variance is just the one taking that value with probability 1.

We assume that Np_m is not integer valued. Then for our distribution Q_m we must have some $k < Np_m$ for which $q_{mk} > 0$, similarly for some $k > Np_m$.

If the only such ks with positive probabilities are $\lfloor Np_m \rfloor$ and $\lceil Np_m \rceil$, then we have the minimal range distribution and our target value $f_m(1-f_m)$ is achieved, as shown in Section 4.2.

In the following, we will distinguish two cases.

Case 1. Suppose that there are $0 \leq k_1 < \lfloor Np_m \rfloor$ with $q_{mk_1} > 0$ and $N \geq k_2 > \lceil Np_m \rceil$ with $q_{mk_1} > 0$. Let $s = \min\{k_1; q_{mk_1} > 0 \text{ and } k_1 < \lfloor Np_m \rfloor\}$, and $l = \max\{k_2; q_{mk_2} > 0 \text{ and } k_2 > \lceil Np_m \rceil\}$. Our distribution thus has range $l - s \geq 2$.

Choose $\alpha = \min(q_{ms}, q_{ml})$. Consider the alternative distribution $Q_m^{/}$ which has the same probabilities over the integers, except that:

$$q_{ms}^{/} = q_{ms} - \alpha, \tag{54}$$

$$q_{ml}^{/} = q_{ml} - \alpha, \tag{55}$$

$$q_{m(s+1)}^{/} = q_{m(s+1)} + \alpha, \tag{56}$$

$$q_{m(l-1)}^{/} = q_{m(l-1)} + \alpha. \tag{57}$$

This distribution has range $l - s - 1$ (or $l - s - 2$ if $q_{ms} = q_{ml}$). The variance of this new distribution is given by

$$\mathrm{Var}(Q'_m) = \mathrm{Var}(Q_m) + \alpha\Big((l - 1 - Np_m)^2 + (s + 1 - Np_m)^2 \tag{58}$$

$$- (l - Np_m)^2 - (s - Np_m)^2\Big) \tag{59}$$

$$= \mathrm{Var}(X_m) - 2\alpha(l - s). \tag{60}$$

It is easy to verify that the above is indeed a probability distribution with mean Np_m. Thus, for any such distribution with range $l - s$ we can find one with a smaller range giving a smaller variance. Thus, no such distribution can achieve the minimum variance.

Case 2. Now consider a distribution where we only have one out of a $k < \lfloor Np_m \rfloor$ and a $k > \lceil Np_m \rceil$ with positive probability. Suppose without loss of generality that we have value $l > \lceil Np_m \rceil$, the largest such value with positive probability (almost identical working yields the equivalent result for only values smaller than $\lfloor Np_m \rfloor$).

The mean of this distribution is Np_m, and we know that the only possible value below the mean is $\lfloor Np_m \rfloor$. Thus as l occurs with probability q_{ml}, $\lfloor Np_m \rfloor$ must occur with probability $q_{m\lfloor Np_m \rfloor}$ satisfying

$$q_{m\lfloor Np_m \rfloor} \geq q_{ml} x \tag{61}$$

where

$$x = \frac{l - Np_m}{Np_m - \lfloor Np_m \rfloor}. \tag{62}$$

Consider the alternative distribution Q'_m which has the same probabilities over the integers, except that:

$$q'_{ml} = 0, \tag{63}$$

$$q'_{m\lceil Np_m \rceil} = q_{m\lceil Np_m \rceil} + (Np_m - \lfloor Np_m \rfloor) q_{ml} (1 + x) \tag{64}$$

$$q'_{m\lfloor Np_m \rfloor} = q_{m\lfloor Np_m \rfloor} - q_{ml} x + (\lceil Np_m \rceil - Np_m) q_{ml} (1 + x) \tag{65}$$

It is easy to verify that the above is indeed a probability distribution with mean Np_m. The range is also clearly smaller than the range $l - \lfloor Np_m \rfloor$ of the original. Moreover,

$$\mathrm{Var}(Q'_m) = \mathrm{Var}(Q_m) - (l - Np_m)^2 - q_{m\lfloor Np_m \rfloor} - q_{ml} x (Np_m - \lfloor Np_m \rfloor)^2 \tag{66}$$

$$+ q_{ml} (1 + x) \Big[(\lceil Np_m \rceil - Np_m)^2 (Np_m - \lfloor Np_m \rfloor) \tag{67}$$

$$+ (Np_m - \lfloor Np_m \rfloor)^2 (\lceil Np_m \rceil - Np_m)\Big] \tag{68}$$

$$= \mathrm{Var}(Q_m) - q_{ml}(Np_m - \lfloor Np_m \rfloor)(\lceil Np_m \rceil - Np_m). \tag{69}$$

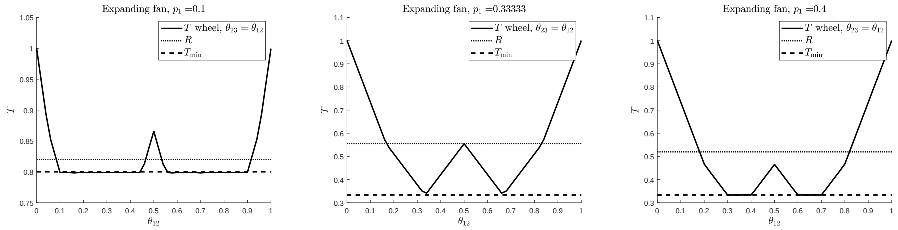

Figure 8. T as a function of angles between spikes. $N = 3, M = 2$. θ_{12} varies, $\theta_{23} = \theta_{12}$. Left figure $p_1 = 0.1$, middle figure $p_1 = 1/3$, right figure $p_1 = 0.4$.

Thus for any such distribution we can find an alternative with a smaller range and a smaller variance.

We therefore have that any distribution with a range bigger than $\lceil Np_m \rceil - \lfloor Np_m \rfloor = 1$, does not yield the smallest variance. Thus the distribution that yields the smallest variance is the (unique) one that has range 1.

Appendix B. Achieving the minimal value with the wheel

Suppose that the angle between the spikes on the wheel model from Section 3.1 is $2\pi/N$. It is easy to see that the expected number of individuals on patch P_m is Np_m, and by considering the angles of the first and final individual on any patch we see that the number of individuals on a given patch is either $\lfloor Np_m \rfloor$ or $\lceil Np_m \rceil$, and is exactly the distribution $\widetilde{X}_m$ as described in Section 4.2. Thus we have that the regular wheel precisely achieves the minimal T.

A uniform spread as described above is not the only distribution which yields a minimal T, see Figure 8. Assume we have 3 individuals and two patches that the individuals visit with probabilities p_1 and $p_2 = 1 - p_1$ respectively for $p_1 \leq 1/2$. The minimum value of T is $T_{\min} = 1/3 \max(1, 3 - 6p_1)$.

If $p_1 \leq 1/3$, the minimal value is achieved by the wheel method with the spikes distributed in such a way that all angles are bigger than (or equal to) p_1 (i.e. uniquely with the uniform distribution for $p_1 = 1/3$ but not uniquely if $p_1 < 1/3$). If $1/3 \leq p_1 \leq 1/2$, the minimal T is achieved if all angles are smaller than (or equal to) p_1 for $1/3 \leq p_1 \leq 1/2$.

Acknowledgements

This work was supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 690817.

References

Abrahams, M. V. (1986). Patch choice under perceptual constraints: a cause for departures from an ideal free distribution. *Behavioral Ecology and Sociobiology*, 19(6):409–415.

Ainley, D. (1972). Flocking in Adélie penguins. *Ibis*, 114(3):388–390.

Ainley, D. (2002). *The Adélie penguin: bellwether of climate change*. Columbia University Press.

Armsworth, P. R. and Roughgarden, J. E. (2005). The impact of directed versus random movement on population dynamics and biodiversity patterns. *The American Naturalist*, 165(4):449–465.

Baguette, M., Vansteenwegen, C., Convi, I., and Neve, G. (1998). Sex-biased density-dependent migration in a metapopulation of the butterfly *Proclossiana eunomia*. *Acta oecologica*, 19(1):17–24.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., et al. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the national academy of sciences*, 105(4):1232–1237.

Beauchamp, G. (1998). The effect of group size on mean food intake rate in birds. *Biological Reviews*, 73(4):449–472.

Beauchamp, G. (2002). Higher-level evolution of intraspecific flock-feeding in birds. *Behavioral Ecology and Sociobiology*, 51(5):480–487.

Beauchamp, G. (2008). What is the magnitude of the group-size effect on vigilance? *Behavioral Ecology*, 19(6):1361–1368.

Belthoff, J. R. and Gauthreaux Jr, S. A. (1991). Partial migration and differential winter distribution of House Finches in the eastern United States. *The Condor*, 93(2):374–382.

Broom, M., Lafaye, C., Pattni, K., and Rychtář, J. (2015). A study of the dynamics of multi-player games on small networks using territorial interactions. *Journal of Mathematical Biology*, 71(6-7):1551–1574.

Broom, M. and Rychtář, J. (2012). A general framework for analysing multiplayer games in networks using territorial interactions as a case study. *Journal of Theoretical Biology*, 302:70–80.

Broom, M. and Rychtář, J. (2018). Ideal cost-free distributions in structured populations for general payoff functions. *Dynamic Games and Applications*, 8(1):79–92.

Buhl, J., Sumpter, D. J., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R., and Simpson, S. J. (2006). From disorder to order in marching locusts. *Science*, 312(5778):1402–1406.

Bullock, J. M., Kenward, R. E., and Hails, R. S. (2002). *Dispersal ecology: 42nd symposium of the British ecological society*, volume 42. Cambridge University Press.

Chapman, C. A. (1990). Association patterns of spider monkeys: the influence of ecology and sex on social organization. *Behavioral Ecology and Sociobiology*, 26(6):409–414.

Collett, M., Despland, E., Simpson, S. J., and Krakauer, D. C. (1998). Spatial scales of desert locust gregarization. *Proceedings of the National Academy of Sciences*, 95(22):13052–13055.

Collins, E. J., Houston, A. I., and Lang, A. (2002). The ideal free distribution: an analysis of the perceptual limit model. *Evolutionary Ecology Research*, 4(4):471–493.

Cosner, C. (2005). A dynamic model for the ideal-free distribution as a partial differential equation. *Theoretical Population Biology*, 67(2):101–108.

Cressman, R. and Garay, J. (2011). The effects of opportunistic and intentional predators on the herding behavior of prey. *Ecology*, 92(2):432–440.

Cressman, R. and Křivan, V. (2006). Migration dynamics for the ideal free distribution. *The American Naturalist*, 168(3):384–397.

Cressman, R., Křivan, V., and Garay, J. (2004). Ideal free distributions, evolutionary games, and population dynamics in multiple-species environments. *The American Naturalist*, 164(4):473–489.

Cunningham, J. M., Calhoun, A. J., and Glanz, W. E. (2006). Patterns of beaver colonization and wetland change in Acadia National Park. *Northeastern Naturalist*, 13(4):583–597.

Danielson, B. J. and Gaines, M. S. (1987). The influences of conspecific and heterospecific residents on colonization. *Ecology*, 68(6):1778–1784.

Despland, E. and Simpson, S. J. (2005). Food choices of solitarious and gregarious locusts reflect cryptic and aposematic antipredator strategies. *Animal Behaviour*, 69(2):471–479.

Dingle, H. (2006). Animal migration: is there a common migratory syndrome? *Journal of Ornithology*, 147(2):212–220.

Dingle, H. (2014). *Migration: the biology of life on the move*. Oxford University Press, USA.

Dingle, H. and Drake, V. A. (2007). What is migration? *Bioscience*, 57(2):113–121.

Emlen, J. T. (1952). Flocking behavior in birds. *The Auk*, 69(2):160–170.

Erovenko, I. V., Bauer, J., Broom, M., Pattni, K., and Rychtář, J. (preprint, 2019). The effect of network topology on optimal exploration strategies and the evolution of cooperation in a mobile population.

Felleman, F. L. (1986). *Feeding ecology of the killer whale (Orcinus orca).* PhD thesis, University of Washington.

Ford, J. R. and Swearer, S. E. (2013). Two's company, three's a crowd: Food and shelter limitation outweigh the benefits of group living in a shoaling fish. *Ecology*, 94(5):1069–1077.

Fretwell, S. D. and Lucas, H. L. (1969). On territorial behavior and other factors influencing habitat distribution of birds. *Acta Biotheoretica*, 19:16–36.

Gadgil, M. (1971). Dispersal: population consequences and evolution. *Ecology*, 52(2):253–261.

Greenberg, R. and Marra, P. P. (2005). *Birds of two worlds: the ecology and evolution of migration.* JHU Press.

Greenwood, P. J. and Harvey, P. H. (1982). The natal and breeding dispersal of birds. *Annual review of ecology and systematics*, 13(1):1–21.

Guttal, V. and Couzin, I. D. (2010). Social interactions, information use, and the evolution of collective migration. *Proceedings of the national academy of sciences*, 107(37):16172–16177.

Hayden, W. and Clifford, H. F. (1974). Seasonal movements of the mayfly *Leptophlebia cupida (Say)* in a brown-water stream of Alberta, Canada. *American Midland Naturalist*, 91:90–102.

Hayward, M. W. and Kerley, G. I. (2005). Prey preferences of the lion (*Panthera leo*). *Journal of Zoology*, 267(3):309–322.

Heimlich-Boran, J. R. (1988). Behavioral ecology of killer whales (*Orcinus orca*) in the Pacific Northwest. *Canadian Journal of Zoology*, 66(3):565–578.

Herrnkind, W. F. and Cummings, W. C. (1964). Single file migrations of the spiny lobster, *Panulirus argus (Latreille).* *Bulletin of Marine Science*, 14(1):123–125.

Hestbeck, J. B. (1982). Population regulation of cyclic mammals: the social fence hypothesis. *Oikos*, pages 157–163.

Hixon, M. A., Pacala, S. W., and Sandin, S. A. (2002). Population regulation: historical context and contemporary challenges of open vs. closed systems. *Ecology*, 83(6):1490–1508.

Houston, A. I. and McNamara, J. M. (1988). The ideal free distribution when competitive abilities differ: an approach based on statistical mechanics. *Animal Behaviour*, 36(1):166–174.

Hugie, D. M. and Grand, T. C. (1998). Movement between patches, unequal competitors and the ideal free distribution. *Evolutionary Ecology*, 12(1):1–19.

Hutto, R. L. (2000). On the importance of en route periods to the conservation of migratory landbirds. *Studies in Avian Biology*, 20:109–114.

Ioannou, C. and Krause, J. (2008). Searching for prey: the effects of group size and number. *Animal Behaviour*, 75(4):1383–1388.

Johnson, N. L. and Kotz, S. (1977). *Urn models and their application; an approach to modern discrete probability theory.* New York, NY (USA) Wiley.

Kappeler, P. M. (2011). Primatological approaches to the study of group coordination. In Boos, M., Kolbe, M., Kappeler, P. M., and Ellwart, T., editors, *Coordination in Human and Primate Groups*, pages 223–228. Springer.

Kennedy, J. (1966). Nervous coordination of instincts. *Cambridge Research*, 2:29–32.

Kim, S.-Y., Torres, R., and Drummond, H. (2009). Simultaneous positive and negative density-dependent dispersal in a colonial bird species. *Ecology*, 90(1):230–239.

King, A. J. (2010). Follow me! I'm a leader if you do; I'm a failed initiator if you don't. *Behavioural Processes*, 84(3):671–674.

Krause, J. and Ruxton, G. D. (2002). *Living in groups.* Oxford University Press.

Křivan, V., Cressman, R., and Schneider, C. (2008). The ideal free distribution: a review and synthesis of the game-theoretic perspective. *Theoretical Population Biology*, 73(3):403–425.

Křivan, V. and Jana, D. (2015). Effects of animal dispersal on harvesting with protected areas. *Journal of Theoretical Biology*, 364:131–138.

Kruuk, H. (1972). *The spotted hyena: a study of predation and social behavior.* University of Chicago Press.

Kuussaari, M., Saccheri, I., Camara, M., and Hanski, I. (1998). Allee effect and population dynamics in the Glanville fritillary butterfly. *Oikos*, pages 384–392.

Leshem, Y. and Yom-Tov, Y. (1998). Routes of migrating soaring birds. *Ibis*, 140(1):41–52.

Mariani, P., Křivan, V., MacKenzie, B. R., and Mullon, C. (2016). The migration game in habitat network: the case of tuna. *Theoretical Ecology*, 9(2):219–232.

Matthysen, E. (2005). Density-dependent dispersal in birds and mammals. *Ecography*, 28(3):403–416.

McPeek, M. A. and Holt, R. D. (1992). The evolution of dispersal in spatially and temporally varying environments. *The American Naturalist*, 140(6):1010–1027.

Mech, L. (1981). *The wolf: Ecology and behavior of an endangered species 1st ed.* University of Minnesota Press.

Mehlman, D. W., Mabey, S. E., Ewert, D. N., Duncan, C., Abel, B., Cimprich, D., Sutter, R. D., and Woodrey, M. (2005). Conserving stopover sites for forest-dwelling migratory landbirds. *The Auk*, 122(4):1281–1290.

Metz, J. and Gyllenberg, M. (2001). How should we define fitness in structured metapopulation models? Including an application to the calculation of evolutionarily stable dispersal strategies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1466):499–508.

Moore, F., Smith, R., and Sandberg, R. (2005). Birds of two worlds: the ecology and evolution of migration. *Stopover ecology of intercontinental migrants: en route problems and consequences for reproductive performance*, pages 251–261.

Moore, F. R. and Yong, W. (1991). Evidence of food-based competition among passerine migrants during stopover. *Behavioral Ecology and Sociobiology*, 28(2):85–90.

Morris, D. W. (2003). Toward an ecological synthesis: a case for habitat selection. *Oecologia*, 136(1):1–13.

Mueller, T., O'Hara, R. B., Converse, S. J., Urbanek, R. P., and Fagan, W. F. (2013). Social learning of migratory performance. *Science*, 341(6149):999–1002.

Murdoch, W. W. (1994). Population regulation in theory and practice. *Ecology*, 75(2):271–287.

O'Brien, D. (1987). Direct observations of the behavior of *Euphausia superba* and *Euphausia crystallorophias (Crustacea: Euphausiacea)* under pack ice during the Antarctic spring of 1985. *Journal of Crustacean Biology*, 7(3):437–448.

O'Neal, B. J., Stafford, J. D., and Larkin, R. P. (2012). Stopover duration of fall-migrating dabbling ducks. *The Journal of Wildlife Management*, 76(2):285–293.

Parker, G. A. (1978). Searching for mates. In Krebs, J. R. and Davies, N. B., editors, *Behavioural ecology: an evolutionary approach*, pages 214–244. Oxford, UK: Blackwell Scientific.

Parrish, J. (1999). Complexity, pattern, and evolutionary trade-off in animal aggregation. *Science*, 284(99):101.

Parrish, J. K. (1989). Layering with depth in a heterospecific fish aggregation. *Environmental Biology of Fishes*, 26(2):79–85.

Parrish, J. K., Viscido, S. V., and Grunbaum, D. (2002). Self-organized fish schools: an examination of emergent properties. *The biological bulletin*, 202(3):296–305.

Pattni, K., Broom, M., and Rychtář, J. (2017). Evolutionary dynamics and the evolution of multiplayer cooperation in a subdivided population. *Journal of Theoretical Biology*, 429:105–115.

Pattni, K., Broom, M., and Rychtář, J. (2018). Evolving multiplayer networks: Modelling the evolution of cooperation in a mobile population. *Discrete & Continuous Dynamical Systems-B*, 23(5):1975–2004.

Pennycuick, C. (1972). Soaring behaviour and performance of some East African birds, observed from a motor-glider. *Ibis*, 114(2):178–218.

Petit, O. and Bon, R. (2010). Decision-making processes: the case of collective movements. *Behavioural Processes*, 84(3):635–647.

Pitcher, T., Magurran, A., and Winfield, I. (1982). Fish in larger shoals find food faster. *Behavioral Ecology and Sociobiology*, 10(2):149–151.

Poethke, H. J. and Hovestadt, T. (2002). Evolution of density–and patch–size–dependent dispersal rates. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1491):637–645.

Pyritz, L. W., Kappeler, P. M., and Fichtel, C. (2011). Coordination of group movements in wild red-fronted lemurs (*Eulemur rufifrons*): processes and influence of ecological and reproductive seasonality. *International Journal of Primatology*, 32(6):1325–1347.

Reding, I., Kelley, M., Rowell, J. T., and Rychtář, J. (2016). A continuous ideal free distribution approach to the dynamics of selfish, cooperative and kleptoparasitic populations. *Royal Society Open Science*, 3(11):160788.

Resetarits, W., Binckley, C. A., and Chalcraft, D. R. (2005). Habitat selection, species interactions, and processes of community assembly in complex landscapes. *Metacommunities: spatial dynamics and ecological communities. University of Chicago Press, Chicago*, pages 374–398.

Resetarits Jr, W. J. and Binckley, C. A. (2009). Spatial contagion of predation risk affects colonization dynamics in experimental aquatic landscapes. *Ecology*, 90(4):869–876.

Rogers, T. and Bryden, M. (1995). Predation of Adélie penguins (*Pygoscelis adeliae*) by leopard seals (*Hydrurga leptonyx*) in Prydz Bay, Antarctica. *Canadian Journal of Zoology*, 73(5):1001–1004.

Rowell, J. T. (2009). The limitation of species range: a consequence of searching along resource gradients. *Theoretical population biology*, 75(2-3):216–227.

Ruxton, G. D. and Humphries, S. (2003). Non-ifd movements: reflections on past work and prospects for future developments. *Evolutionary Ecology Research*, 5(1):155–157.

Scheel, D. (1993). Profitability, encounter rates, and prey choice of african lions. *Behavioral Ecology*, 4(1):90–97.

Scheiber, I. B., Weiß, B. M., Kotrschal, K., and Hemetsberger, J. (2013). *The social life of Greylag Geese.* Cambridge University Press.

Schneider, D. W. and Frost, T. M. (1986). Massive upstream migrations by a tropical freshwater neritid snail. *Hydrobiologia*, 137(2):153–157.

Sillett, T. S. and Holmes, R. T. (2002). Variation in survivorship of a migratory songbird throughout its annual cycle. *Journal of Animal Ecology*, 71(2):296–308.

Simpson, S., Despland, E., Hägele, B., and Dodgson, T. (2001). Gregarious behavior in desert locusts is evoked by touching their back legs. *Proceedings of the National Academy of Sciences*, 98(7):3895–3897.

Stamps, J. (1991). The effect of conspecifics on habitat selection in territorial species. *Behavioral Ecology and Sociobiology*, 28(1):29–36.

Stander, P. E. (1992). Cooperative hunting in lions: the role of the individual. *Behavioral Ecology and Sociobiology*, 29(6):445–454.

Street, G. M., Erovenko, I. V., and Rowell, J. T. (2018). Dynamical facilitation of the ideal free distribution in nonideal populations. *Ecology and Evolution*, 8(5):2471–2481.

Sutherland, W. and Parker, G. (1992). The relationship between continuous input and interference models of ideal free distributions with unequal competitors. *Animal Behaviour*, 44:345–355.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Sword, G. A. and Dopman, E. B. (1999). Developmental specialization and geographic structure of host plant use in a polyphagous grasshopper, *Schistocerca emarginata* (=*lineata*)(Orthoptera: Acrididae). *Oecologia*, 120(3):437–445.

Sword, G. A. and Simpson, S. J. (2000). Is there an intraspecific role for density-dependent colour change in the desert locust? *Animal Behaviour*, 59(4):861–870.

Takahashi, A., Sato, K., Nishikawa, J., Watanuki, Y., and Naito, Y. (2004). Synchronous diving behavior of Adélie penguins. *Journal of Ethology*, 22(1):5–11.

Taylor, L. (1986). Synoptic dynamics, migration and the Rothamsted insect survey: Presidential address to the British Ecological Society, December 1984. *Journal of Animal Ecology*, 55(1):1–38.

Todd, F. S. (1988). Weddell seal preys on chinstrap penguin. *The Condor*, 90(1):249–250.

Travis, J. M., Murrell, D. J., and Dytham, C. (1999). The evolution of density–dependent dispersal. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1431):1837–1842.

Tremblay, Y. and Cherel, Y. (1999). Synchronous underwater foraging behavior in penguins. *The Condor*, 101(1):179–185.

Turchin, P. (1999). Population regulation: a synthetic view. *Oikos*, pages 153–159.

Uvarov, B. (1977). *Grasshoppers and locusts. A handbook of general acridology Vol. 2. Behaviour, ecology, biogeography, population dynamics.* Centre for Overseas Pest Research, London, UK.

Van Loon, E., Shamoun-Baranes, J., Bouten, W., and Davis, S. (2011). Understanding soaring bird migration through interactions and decisions at the individual level. *Journal of Theoretical Biology*, 270(1):112–126.

van Noordwijk, A. J., Pulido, F., Helm, B., Coppack, T., Delingat, J., Dingle, H., Hedenström, A., van der Jeugd, H., Marchetti, C., Nilsson, A., et al. (2006). A framework for the study of genetic variation in migratory behaviour. *Journal of Ornithology*, 147(2):221–233.

Voelkl, B. and Fritz, J. (2017). Relation between travel strategy and social organization of migrating birds with special consideration of formation flight in the northern bald ibis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1727):20160235.

Ward, C. R., Gobet, F., and Kendall, G. (2001). Evolving collective behavior in an artificial ecology. *Artificial life*, 7(2):191–209.

Watts, D. P. (1994). The influence of male mating tactics on habitat use in mountain gorillas (*Gorilla gorilla beringei*). *Primates*, 35(1):35–47.

White, J. W., Samhouri, J. F., Stier, A. C., Wormald, C. L., Hamilton, S. L., and Sandin, S. A. (2010). Synthesizing mechanisms of density dependence in reef fishes: behavior, habitat configuration, and observational scale. *Ecology*, 91(7):1949–1961.

Whitham, T. G. (1978). Habitat selection by Pemphigus aphids in response to response limitation and competition. *Ecology*, 59(6):1164–1176.

Wilson, R. P., Wilson, M.-P. T., and McQuaid, L. (1986). Group size in foraging African penguins (*Spheniscus demersus*). *Ethology*, 72(4):338–341.

Würsig, B. (1986). Delphinid foraging strategies. In R. J. Schusterman, J. A. Thomas, F. G. W. R. S., editor, *Dolphin cognition and behavior: A comparative approach*, pages 347–359. Taylor and Francis.

Yates, G. E. and Broom, M. (2005). A stochastic model of the distribution of unequal competitors between resource patches. *Journal of Theoretical Biology*, 237(3):227–237.