SUPPORTING INFORMATION

Foraging success under uncertainty: search tradeoffs and optimal space use

Frederic Bartumeus[1,2,3]*, Daniel Campos[4], William S. Ryu[5], Roger Lloret-Cabot[1,2], Vicenç Méndez[4], and Jordi Catalan[2]

[1] Centre for Advanced Studies of Blanes (CEAB-CSIC), Cala Sant Francesc 14, 17300 Girona, Spain.

[2] CREAF, Cerdanyola del Vallès, 08193 Barcelona, Spain.

[3] ICREA, Pg Lluís Companys 23, 08010 Barcelona, Spain.

[4] Grup de Física Estadística. Departament de Física, Universitat Autònoma de Barcelona. Barcelona 08193, Spain.

[5] Department of Physics and the Donnelly Centre, University of Toronto, 60 St George St. Toronto, M5S1A7, Canada.

July 10, 2016

1 T_2: the detection time

In the main manuscript, we have conveniently split the search time into the approaching time, T_1 and the detection time, T_2. The change from one to the other occurs at a distance from the target $x_a \approx \mathcal{O}(v\tau)$ where v is the velocity

*Corresponding author: fbartu@ceab.csic.es

and τ the correlation length, or in other words, a distance on the order of the characteristic flight time (Figure 1). Due to the diffusive nature of the searcher's movement, the approaching time depends explicitly on the coefficient D. However, the detection time is a function of L, v and $p(v)$ (i.e. $T_2 = f(L, v, p(v))$). For the particular case of CRWs we have $T_2 = L/vp(v)$, the time to cover the domain size L ballistically (see Eq. 2 in the main text).

Since the searcher's motion during T_2 is still essentially diffusive, it may be puzzling why T_2 does not depend explicitly on D. To understand why this happens, note the seacher's movement may contain many back-and-forth movements away and towards the target, and that directional persistence plays an ambivalent role. It both reduces the turning rate on arrival, and hence the probability for the searcher to wander around the target, but at the same time increases the characteristic distance of departure (i.e. x_a, Figure 1) from the target, which is $v\tau$ (or should be proportional to it).

At short-scales, the movement of the searcher towards the target can also be interpreted as an approaching time but averaged over positions $> x_a$. Expressed in mathematical form, this is

$$\langle T_2 \rangle = \int_0^L \rho(x_a) \frac{x_a(L - x_a)}{2D} dL \tag{1}$$

where $\rho(x_a)$ represents the probability distribution of the distance x_a. Now, since we are assuming that the domain size is much larger than the typical flight distance, $L \gg x_a$ we can simplify the integral to obtain

$$\langle T_2 \rangle = \int_0^L \rho(x_a) \frac{x_a L}{2D} dL = \frac{L}{2D} \langle x_a \rangle \sim L/v \tag{2}$$

where in the last identity we have used that the average value of x_a is proportional to $v\tau$, as stated above, and the definition of the diffusion coefficient is $D \equiv v^2\tau$. Hence, we observe that the result $T_2 \sim L/v$ is because the distance of departure x_a, at which the detection phase starts, depends explicitly on the motion parameters v and τ. Whenever the searcher moves towards the

target, persistence (which can be attained by increasing either v or τ) facilitates the encounter with the target by decreasing the time spent wandering around it. On the other hand, if the searcher mistakenly moves away from the target persistence increases the characteristic distance of departure x_a. Our calculations show that, if $L \gg x_a$, these two effects compensate for each other in the sense that the resulting time T_2 becomes independent of the diffusion constant D, governed by the characteristic flight time τ. If x_a is small, the departure distances are small but the turning rates upon arrivals become too large. If x_a is large, the departure distances are large but arrivals are less meandrous. The idea that when detection is plausible (T_2 regime) persistence plays opposite roles is valid for any random search process. However, the exact cancellation of such effects only occurs for random walks where the flight time distribution shows a clear-cut scale that is much smaller than the search domain L (e.g. CRWs). The introduction of multiple flight times (or persistence) scales can help solve this conundrum and introduces the possibility of further optimization (Campos *et al.*, 2015).

2 Generalized diffusion coefficient in 2D

In the main manuscript, we provide a general expression for the diffusion coefficient of random walkers

$$D(v, \alpha, \varphi(t)) = \frac{v^2 \left\langle t^2 \right\rangle \left[1 + \left(\frac{2\langle t \rangle^2}{\langle t^2 \rangle} - 1\right) \alpha\right]}{2d \left\langle t \right\rangle (1 - \alpha)} \tag{3}$$

which can be found in various references (e.g. Lovely & Dahlquist, 1975; Dusenbery, 2009). This expression is valid for any distribution of flight times and turn angles, provided they have finite first and second order moments. However, it still has some limitations, as it assumes that (i) the speed of the walkers is fixed, and (ii) turns are instantaneous (so pauses between consecutive flights are not considered).

It is possible to generalize this expression by relaxing these two assumptions.

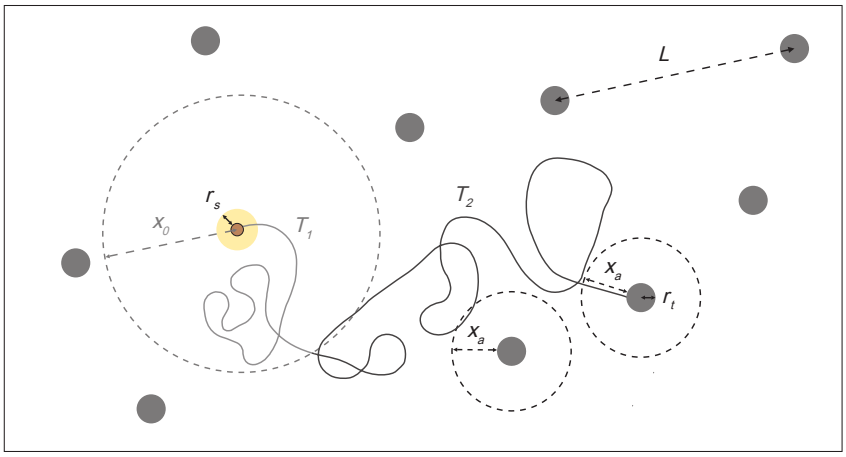

Figure 1: Depiction of the key temporal and spatial scales involved in the computation of the mean-first passage times. Grey filled circles represent targets and the smaller brown, filled circle represents the searcher. r_t and r_s are the size and the perceptual scale of the target and the searcher respectively. L represents the average distance between targets. We depict one single realization of the whole ensemble of search trajectories, divided into two relevant temporal scales T_1 and T_2. T_1 is the mean time necessary to leave an empty area and approach a target. T_1 is a function of the spatial scale x_0, which delimits the distance (grey dashed-circle area in two dimensions) that the searchers need to cross to reach the closest target, in other words, the minimal distance required to leave an empty area. T_2 is the mean time needed to detect a target once the searcher trajectories are arbitrarily close to any target such that an *average* detection is possible. T_2 is a function of the spatial scale x_a, the characteristic distance of departure/arrival from/to a target (black dashed-circle in two dimensions), which is proportional to the characteristic flight time (or persistence) $v\tau$, where v is the velocity and τ is the correlation length. When detection is plausible (i.e. T_2) persistence plays opposite roles. If x_a is large, the departure distances are large but arrivals are less meandrous. The opposite is true for small x_as.

The derivation, which is built on the foundations of d-dimensional Continuous-Time Random Walks is is lengthy, and will not be included here but will be published in a more technical article. However, we provide here the general result because it is needed to complete the discussion in Section 3.

The diffusion coefficient, in the more general case, is

$$D(v,\alpha,\varphi(t)) = \frac{\left\langle t^2 \right\rangle \left[\left\langle v^2 \right\rangle + \left(\frac{2\langle t \rangle^2 \langle v \rangle^2}{\langle t^2 \rangle} - \left\langle v^2 \right\rangle \right) \alpha \right]}{2d \left(\langle t \rangle + \langle t_p \rangle \right) (1-\alpha)} \tag{4}$$

where $\langle t_p \rangle$ is the mean time of the pause distribution (the mean time the walker waits between the end of one flight and the beginning of the next one) and $\langle v \rangle$, $\left\langle v^2 \right\rangle$ are the first and second order moments of the speed distribution, as we now consider that flight speeds are random and follow a given probability distribution.

From the expression (4) in addition to computing the diffusion coefficient of the *enhanced* or the *composite* case discussed in the main text, but one can also consider much more general trajectories for which speeds, pause times, flight time, and turn angles are characterized by their corresponding probability distributions.

3 Enhanced and composite diffusion

In this Section, we provide the details for the derivation of the diffusion coefficient in the *enhanced* and the *composite* cases discussed in the main text.

3.1 Enhanced diffusion

One way to generate multi-scale search patterns is through episodes of long-lasting directional persistence, the so-called *relocations*. At a statistical level, relocations facilitate the emergence of heavy-tailed distributions of flight times and/or flight distances, which in turn, yields *enhanced diffusion*, and determine the scaling $MSD \approx t^{\gamma}$ for the mean square displacement (MSD) over a range

of scales (Raposo *et al.*, 2011; Bartumeus *et al.*, 2014).

Truncated Lévy flights, which are governed by power-law (heavy-tailed) flight distributions with an upper and a lower cutoff (representing intrinsic biological limitations) have become a paradigm in search theory, borrowed from statistical mechanics, to explore these ideas. Since truncated power-law distributions have finite moments, a diffusion coefficient can also be formally defined for truncated Lévy flights.

While the ubiquity of Lévy patterns in animal movement has been largely questioned, in particular, regarding the statistical procedures used to fit power-laws (Edwards *et al.*, 2007; Edwards, 2011; Petrovskii *et al.*, 2011; Jansen *et al.*, 2012; Reynolds, 2012), it is evident that (i) long relocations commonly arise in animal trajectories, often leading to slower-than-exponential decays in flights distributions, and (ii) the landscape features and external cues are not enough to explain these patterns, since they can be also observed under homogeneous or otherwise simple environments (Bazazi *et al.*, 2012; Campos *et al.*, 2014; Salvador *et al.*, 2014). In addition, recent works (Bartumeus *et al.*, 2014) have shown that heavy-tailed distributions of directional change different than Lévy can generate similar statistical signatures, achieving similar search efficiency as Lévy flight models.

For a truncated Lévy flight characterised by a flight time distribution

$$\varphi_{enh}(t) = \frac{\mu}{t_{mmin}^{-\mu} - t_{max}^{-\mu}} t^{-1-\mu},$$

with μ positive, the computation of the diffusion coefficient (according to Eq. 3 in the main text) requires the determination of the first and second moments of this distribution. By definition these are

$$\langle t \rangle = \int_{t_{min}}^{t_{max}} \varphi_{enh}(t) dt = \begin{cases} \frac{\mu}{\mu - 1} \frac{t_{min}^{1-\mu} - t_{max}^{1-\mu}}{t_{min}^{-\mu} - t_{max}^{-\mu}} & \mu \neq 1 \\ \frac{\mu}{t_{min}^{-\mu} - t_{max}^{-\mu}} \log \frac{t_{max}}{t_{min}} & \mu = 1 \end{cases}$$

$$\langle t^2 \rangle = \int_{t_{min}}^{t_{max}} \varphi_{enh}(t) dt = \begin{cases} \frac{\mu}{\mu-2} \frac{t_{min}^{2-\mu} - t_{max}^{2-\mu}}{t_{min}^{-\mu} - t_{max}^{-\mu}} & \mu \neq 2 \\ \frac{\mu}{t_{min}^{-\mu} - t_{max}^{-\mu}} \log \frac{t_{max}}{t_{min}} & \mu = 2. \end{cases}$$

Then, if we replace these expressions into the general form of the diffusion coefficient (Equation 4 from the main text)

$$D(v, \alpha, \varphi(t)) = \frac{v^2 \langle t^2 \rangle \left[1 + \left(\frac{2\langle t \rangle^2}{\langle t^2 \rangle} - 1 \right) \alpha \right]}{2d \langle t \rangle (1 - \alpha)}, \tag{5}$$

one obtains the *enhanced* diffusion coefficient (including all possible values of μ)

$$D_{enh} = \begin{cases} \frac{v^2}{2d} \left(\frac{1-\mu}{2-\mu} \frac{t_{max}^{2-\mu} - t_{min}^{2-\mu}}{t_{max}^{1-\mu} - t_{min}^{1-\mu}} + \frac{2\alpha}{1-\alpha} \frac{-\mu}{1-\mu} \frac{t_{max}^{1-\mu} - t_{min}^{1-\mu}}{t_{max}^{-\mu} - t_{min}^{-\mu}} \right) & \mu \neq 1, \mu \neq 2 \\ \frac{v^2}{2d} \left(\frac{1-\mu}{2-\mu} \frac{t_{max}^{2-\mu} - t_{min}^{2-\mu}}{t_{max}^{1-\mu} - t_{min}^{1-\mu}} + \frac{2\alpha\mu}{1-\alpha} \log \frac{t_{max}}{t_{min}} \right) & \mu = 1 \\ \frac{v^2}{2d} \left[(1-\mu) \log \frac{t_{max}}{t_{min}} + \frac{2\alpha}{1-\alpha} \frac{-\mu}{1-\mu} \frac{t_{max}^{1-\mu} - t_{min}^{1-\mu}}{t_{max}^{-\mu} - t_{min}^{-\mu}} \right] & \mu = 2. \end{cases} \tag{6}$$

3.2 Composite diffusion

Another way to generate a multi-scale search pattern is by utilizing different characteristic scales. Given these scales, diffusion coefficients can be computed from composite Brownian motion random walks.

Composite Brownian motion is often interpreted as the result of the behavioural reactions to landscape features and cues (Schick *et al.*, 2008; Fronhofer *et al.*, 2013; Benhamou, 2014). Hence, the pattern emerges from the interaction with the landscape (which already display multi-scale properties), and is not necessarily generated intrinsically by the organism (Petrovskii *et al.*, 2011; Benhamou, 2014). Recent empirical evidence, however, suggests that such composite motion may not be completely coupled to landscape features but rather it may be internally generated (de Jager *et al.*, 2011; Jansen *et al.*, 2012; de Jager *et al.*, 2012, 2014). Importantly, if the composite diffusion is generated by a specific set of characteristic scales, it would resemble a Lévy walk (Reynolds, 2014). Current empirical research (de Jager *et al.*, 2011; Jansen *et al.*, 2012;

de Jager *et al.*, 2014) suggests that this could be the case.

Diffusion coefficients can also be computed for composite Brownian motion random walks. The derivation of the diffusion coefficient follows similar arguments to those for the *enhanced* case. In particular, we implement the idea of a multi-scale motion pattern by introducing a hyper-exponential flight time distribution $\varphi_{comp}(t)$; while this is not the only way to address composite random walks, it is certainly the most natural one. For the simplest case with only two scales $\langle t_1 \rangle$ and $\langle t_2 \rangle$ whose corresponding weights are w and $1-w$ one has then $\varphi_{comp}(t) = \frac{w}{\langle t_1 \rangle} e^{-t/\langle t_1 \rangle} + \frac{(1-w)}{\langle t_2 \rangle} e^{-t/\langle t_2 \rangle}$. The diffusion coefficient computed from Eq. 3 has the form

$$D_{comp} = D(v, 0, \varphi_{comp}(t)) = \frac{v^2 \left(w\langle t_1 \rangle^2 + (1-w)\langle t_2 \rangle^2 \right)}{d \left(w\langle t_1 \rangle + (1-w)\langle t_2 \rangle \right)}. \tag{7}$$

More generally, for N different movement scales $\langle t_1 \rangle, \ldots, \langle t_N \rangle$ with weights $w_1, w_2, \ldots, w_N$ one has $\varphi_{comp}(t) = \sum_{i=1}^{N} \frac{w_i}{\langle t_i \rangle} e^{-t/\langle t_i \rangle}$, which gives

$$D_{comp} = \frac{v^2 \sum_{i=1}^{N} w_i \langle t_i \rangle^2}{d \sum_{i=1}^{N} w_i \langle t_i \rangle}. \tag{8}$$

4 The speed-perception tradeoff in 1D

The speed-perception tradeoff depends on how the ability to detect nearby targets varies with speed (Figure 2). The impact of the speed on search efficiency, mediated by the speed-perception tradeoff, can be more clearly shown by Monte Carlo simulations of random searchers with different flight time distributions, moving in one-dimension, under two limiting search regimes (symmetric and asymmetric).

The detection ability is taken into account by assuming that every flight that passes within a distance R to the target location has a detection probability of $p(v) = e^{-\gamma v}$, where $\gamma > 0$ represents a detection parameter, and R is the effective detection distance or the target size. This detection mechanism penalizes passages over the target at high speeds, which will have a reduced detection

rate. Note that the choice for $p(v)$ does not necessarily correspond to a particular type of perceptual response (e.g. visual) for any particular organism or situation, but it is simply used here as a general function to cover all possible levels of detection ability from $\gamma = 0$ (perfect detection upon encounter) to γ large (poor detection even at relatively low speeds).

The speed-perception tradeoff requires finding an optimal cruising speed, that is, the maximum speed possible, accounting for energetic considerations (Pyke, 1981), with a minimum of perception loss (O'Brien *et al.*, 1990; Campos *et al.*, 2012), or else an optimal combination of fast and slow search modes (Bénichou *et al.*, 2011). Figure 2 shows the MFDT as we increase speed, for different speed-perception values (different values of the parameter γ). Regardless of the distribution of flight times $\varphi(t)$ (exponential or Lévy) and the search regime (asymmetric or symmetric), if detection is perfect (MFDT equivalent to MFPT), the MFDT decreased with increasing speed. However, since speed interferes with perception, an optimal speed minimizing the MFDT emerges.

5 *Caenorhabditis elegans* trajectory analysis

We placed one-by-one 39 individuals (well-fed on a bacteria lawn for several days) onto a bare agar plate of 24.5×24.5 cm at a homogeneous temperature of 21°C. In the bare arena, we tracked the worms at 32 Hz, for about 90 minutes. We reconstructed the worm trajectories based on the coordinates of the centroid of mass. All worms were cultivated under the same temperature conditions as the assay. Individuals were rinsed of *E.coli* by transferring them from OP50 food plates into M9 buffer (same inorganic ion concentration as M9 assay plates) and letting them swim for 1 min. Individual worms were transferred from the M9 buffer to the centre of the assay plate. The first 3 minutes the behaviour of the animal was affected by manipulation (gentle translocation from one agar plate to the other) and acclimatization to a new environment. We began the data analysis after the worms had 5 minutes to fully recover to basal behaviour.

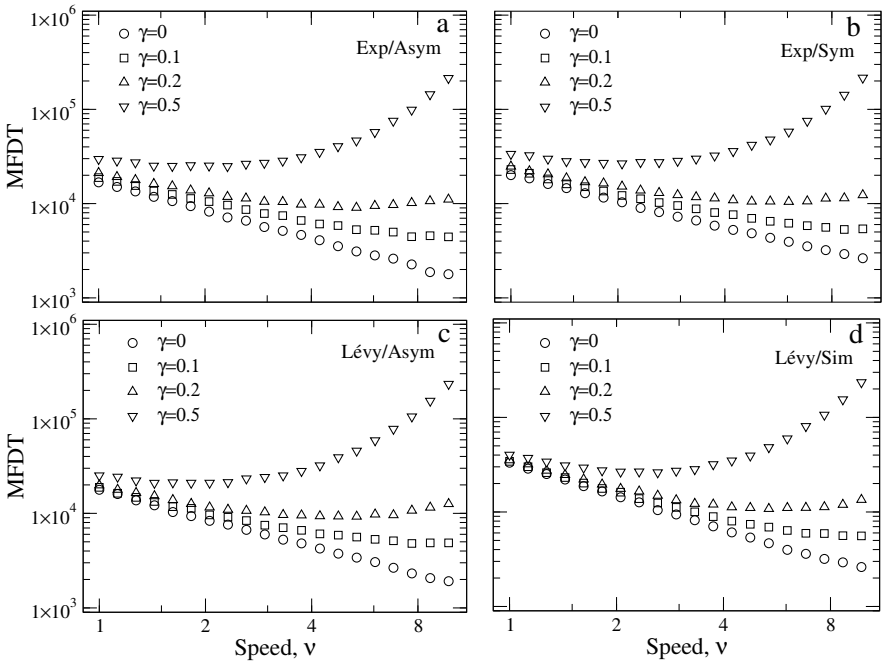

Figure 2: Mean First-Detection Time (MFDT) as a function of movement speed and for different values of the detection parameter γ. The larger the γ the more difficult it is for the searcher to detect the target. Correlated random walk (CRW): (a) asymmetric, (b) symmetric search conditions; Enhanced diffusion (D_{enh}, truncated Lévy): (c) asymmetric, (d) symmetric search conditions. If detection is almost perfect (small values of γ), the larger the speed the smaller the MFDT. However, as detection probability decreases (large γ values) the speed increases the MFDT and an intermediate speed emerges as optimal. This is true for both models (exponential and truncated Lévy) and search initial conditions (asymmetric and symmetric).

5.1 Movement variables used as input features

N_i is the neighbourhood of any location i in the trajectory, given as a subset of successive locations centred on i. l, r is the leftmost and rightmost locations in any neighbourhood set, and d_{ij} is the Euclidean distance between any two locations i, j. We defined three spatial measures averaged over 5 minute windows: Straightness/Sinuosity Index, the Net Displacement, and the Mean Velocity.

- **Straightness Index**

$$S_i = \frac{1}{|N_i|} \sum_{j \in N_i} d_{ij}. \tag{9}$$

  This is an inverse measure of the spatial aggregation of neighbouring locations, and characterizes the intensity of the local search.

- **Net displacement**

$$D_i = d_{l,r}^{(i)}. \tag{10}$$

  This is the mean *net displacement* over all locations in the neighbourhood of i, where $d_{l,r}^{(j)}$ is the net displacement for each location $j \in N_i$ including i itself. This measures the tendency of the individual to move to a different location. Dividing this by the time span of our observation window (300 seconds), gives an *effective velocity*.

$$V_i = \frac{d_{l,r}^{(i)}}{300}. \tag{11}$$

- **Mean travel/speed**

$$T_i = \frac{1}{|N_i|} \sum_{j \in N_i} d_{j,j+1}. \tag{12}$$

  This is the mean *displacement or travelled distance*, where N is the number of neighbouring locations within the observation window, and scales the

measure to an order of magnitude that is similar to the other measures defined. This measures the individuals average speed. Dividing this by the average time span between neighbouring locations (≈ 3 seconds), gives the mean velocity within the observation window.

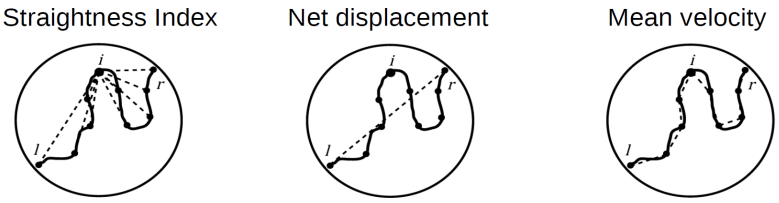

Figure 3: Depiction of the computation of the movement variables used in the behavioural mode analysis of *C.elegans* trajectories.

5.2 Behavioural modes classification

Using these three variables as input features (Figure 3), we constructed a behavioural landscape and partitioned it following the procedure described in Berman *et al.* (2014), which involves the use of a **t-Stochastic Neighbouring Embedding** algorithm (van der Maaten & Hinton, 2008; Berman *et al.*, 2014) along with some post processing.

The t-Stochastic Neighbouring Embedding (t-SNE) is a dimensionality reduction (embedding) algorithm to visualize potential clustering structures existing in the data sets. The principle of *embedding* is to *preserve the similarities* between data points. Similarities do not necessarily need to be expressed as Euclidean distances but usually are related. In other words, data points with high similarities in the high-dimensional space are mapped closely in the low-dimensional space while data points with low similarities are mapped separately. The t-SNE is computationally expensive but can be implemented in a simplified form (Barnes-Hut approximations van der Maaten & Hinton (2008)) of the order $N \, log \, N$ suitable for large, real-world, high-dimensional data sets.

The output of the t-SNE algorithm depends on a basic parameter called perplexity P (comparable with the number of nearest neighbours that is em-

ployed in many manifold learners) that needs to be explored, and also on a stop criteria of a maximum number of iterations or a minimum step-improvement. Furthermore, it is advisable to perform some pre-processing of the input data (*i.e.* feature selection, filtering, PCA, standardisation).

To obtain a final unsupervised set of behavioural modes, the t-SNE output needs some post-processing (Figure 4). First we need to compute a **kernel density estimation** (KDE) upon the embedded space to generate a continuous behavioural landscape whose ruggedness/smoothness is modulated by a parameter *H*. In this way, one can detect areas with high concentration of data points (peaks) at different scales. Second, we compute a **watershed transformation** (WSHD) using a specified connectivity *CONN* to split the behavioural landscape into discrete clusters or polygons. This post-processing adds two new parameters to the whole analysis (*H* and *CONN*).

This procedure allows one to describe movement behavioural states in a principled way and as a hierarchical set of modules (Berman *et al.*, 2014). Based on the input features described (Figure 3), we observed three behavioural clusters related to exploitation, exploration, and relocation behaviours. Figure 5 in the MS and Figure 4 show the closest behavioural landscape to the *average* behavioural landscape obtained after exploring the parameter space over *H*s and *P*s, and running hundreds of seeds over a subset of these values. Further methodical research is needed to better systematize t-SNE behavioural analyses. Overall, we find this unsupervised procedure to be a good way to assess the presence of movement behavioural modes with as few assumptions as possible. However, this does not mean that the methodology is completely independent of the parameterization but the robustness of the results can be explored across parameters. The actual behavioural landscape and its partition is also dependent on the input features. Hence, radically changing the behavioural descriptors may change the number and type of behavioural modes found.

We also explored behavioural segmentation by means of **Hidden Markov Models (HMM)**. In this case, we pre-assumed the presence of three states,

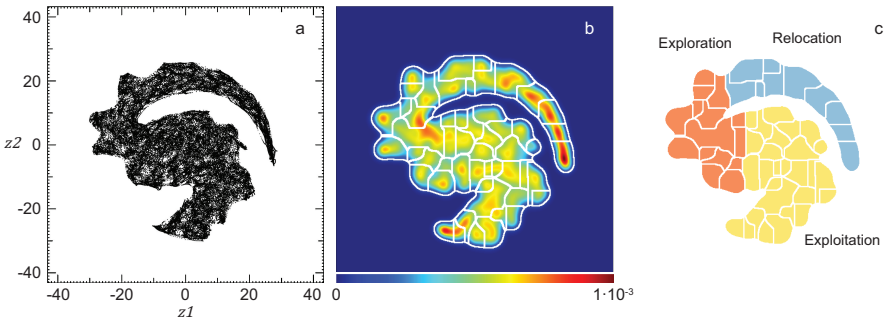

Figure 4: Quantitative analysis of the worm *Caenorhabditis elegans* long-term (90 minutes) search movement in a bare arena. Analysis performed over a total of 69,035 data points. (a) Stochastic Neighbouring Embedding (t-SNE, P=1020), (b) Kernel Density Estimation (KDE, H=19), and (c) Watershed (WSHD, CONN=8) algorithm outputs. Based on three trajectory variables (i.e. Straightness Index, Net Displacement, and Mean Velocity) averaged over 5 min windows we obtained a behavioural landscape partition in three large modules representing: exploitation, exploration, and relocation movement behaviour.

and based on the model fit we obtained qualitatively similar results (note that states' prevalences and mean square displacement scaling exponents slightly differ between the t-SNE and the HMM approach). The results in Figure 5 suggest some degree of statistical coherence and robustness in our behavioural analysis.

5.3 Significance and robustness of the 3-state case

Any characterization of a trajectory into behavioural modes needs specific input features (variables) and parameterization, therefore, there is always some degree of subjectivity. In addition, here we are trying to infer behaviour from the animal's trajectory and it is unclear how movement variables are coupled to the intrinsic (i.e. *hidden*) behavioural states we are searching for. Most likely, the movement-related modes do not represent true behavioural states and so the most we can do is characterize them statistically and hope that their signficance and robustness represent something close to an instrinsic behavioural state of the animal.

In this section, we assess the significance and robustness of the three move-

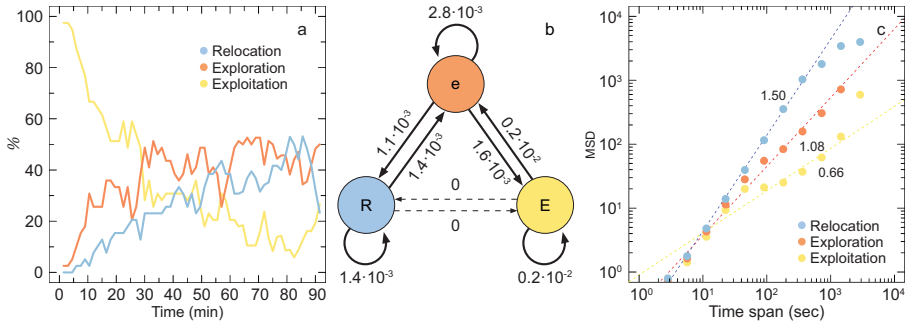

Figure 5: Behavioural segmentation based on a 3-state Hidden Markov Model. (a) HMM states' prevalences (probability of being in a given state) through time. (b) Transition probabilities among states of the fitted model (E=exploitation, e=exploration, and R=relocation). (c) Logartihmic binning plot of the mean square displacement (MSD) with time. The three modes identified show distinct long-term diffusive properties, ranging from subdiffusion to superdiffusion.

ment modes observed by performing a broader analysis using both the Stochastic Neighbouring Embedding (t-SNE) and the Hidden Markov Modelling (HMM) approach. Both methods can show a different number of behavioural states depending on the parameterization. In the *t-SNE-KDE-WSHD* procedure, the effect of the Kernel Density Estimation (KDE) parameter (H) is similar to imposing a number of states when using HMMs. The difference is that the number of states emerge when coarse-graining the behavioural landscape generated from the t-SNE space (i.e. the smaller the H parameter the larger the number of states). Therefore, pooling the whole data set, and by fixing the rest of the parameters, we explored a wide range of values of H (from coarse to high-resolution landscapes), and obtained a different series of states' labels corresponding to partitions into a different number of states. In particular, we covered the following values of $H = \{22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12\}$, leading to the following number of states $S = \{2, 3, 3, 3, 3, 5, 5, 7, 7, 8, 8\}$, respectively. In addition, we ran HMMs and obtained the states' labels under the assumption of different underlying number of states (from 2 to 9).

For the case of the HMMs, we were able to estimate the log-likelihood, the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) of the models, applied to the full dataset (Burnham & Anderson, 2002).

All of them showed a monotonic behaviour (log-likelihood increasing, and AIC and BIC decreasing) with a very slight saturation effect as the number of states increased. This type of profiling is a common effect when assuming Markov chain conditions upon a dataset that does not truly fulfill the Markov assumptions, and cannot be taken as a clear indication to choose any particular model, unless a strong saturation or threshold-like effect is observed (see for example Dean *et al.* (2013)).

Linear discriminant and leave-one-out strategy

As the t-SNE is a computational procedure without an underlying behavioural model, classic model selection based on information criteria (AIC, BIC) does not apply. Because of this we used a different approach based on linear discriminant analysis (LDA). LDA is a method that searches for a linear discriminant (LD: a linear combination of the input features) that separates the classes by maximizing the ratio of the intra-cluster variance with respect to the total variance. As LDA is a supervised method, for each individual we can use the state's labels given by the t-SNE method with different H's or the ones given by the HMMs as the training set for the LDA. We considered the following hypothesis: the best the partition deduced by the *t-SNE-KDE(H)-WSHD* method should produce the best performance of the LDA in classifying the data points.

We used a leave-one-out strategy, that is, for each number of states (H value or HMM) and for each targeted individual we set up a training and a validation dataset. The training dataset is defined as the whole population of individual trajectories except the trajectory from the targeted individual. The validation dataset is thus formed by one single trajectory, the trajectory from the targeted individual.

For the case of the t-SNE analysis we did the following:

1. Fit a LD to the training set, using the state's labels of the same training set;

2. Predict the state's labels for the validation set (the hold out individual)

using the LD obtained in 1;

3. Compare (confusion matrix) the state's labels obtained for the hold out data (targeted individual) with the two approaches: the *t-SNE+KDE(H)+WSHD* procedure and the LDA prediction.

The comparisons between the results of the *t-SNE-KDE(H)-WSHD* procedure and the LDA prediction, per each individual and number of states (H value), were summarized with the F-measure statistic of the confusion matrix. For each H (or number of states) we plotted the individual F-measure values and added an average line (Figure 6, left). First of all, note that as we change the parameter H (from 22 to 12), the number of states ranges from 2 to 8, with the 3-state case the most stable/robust across H's (note that in this case the number of states emerge from the topology of the behavioural landscape itself at different smoothness levels). In addition, the capacity of the LDA to discriminate among the behavioural states identified peaks at the 3-state cases and then decreases (Figure 6, left). The LD fit is only slightly worse for the 2-state case compared to the 3-state case, but the 2-state case is quite unstable and rapidly transition to the 3-state case as we move along H (Figure 6, left). Overall, the 3-state case looks like a better compromise between the statistical significance, measured here as the discrimination capacity of the LD, and the robustness, measured here as the number of times the t-SNE space shows 3 clusters as we move from smooth to rugged landscapes.

For the HMM analysis we did the same exercise as with the t-SNE procedure but with HMMs diverging in a number of states (from 2 to 9). We proceeded with the *leave-one-out* strategy as the one used for the t-SNE procedure but with some differences. In this case, for each targeted trajectory and number of states (2 to 9) we performed the following steps:

1. We learned a HMM using the training set and we decoded the sequence of states (the Viterbi algorithm sequence) of this training set.

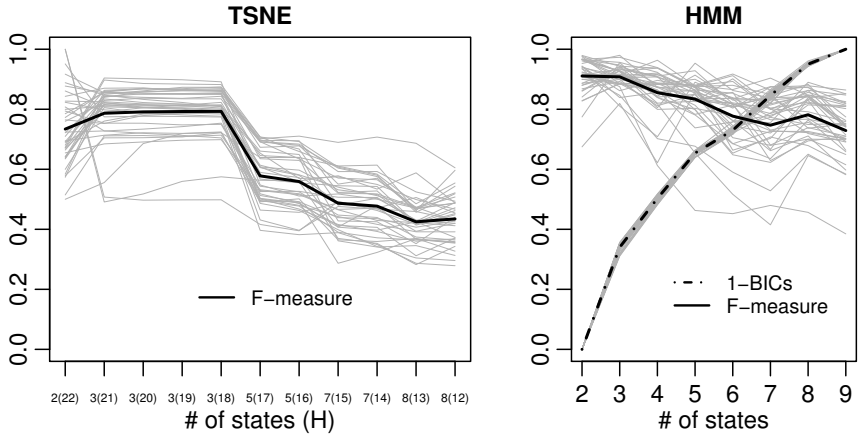

Figure 6: Linear discriminant and leave-one-out analysis. F-measure curves comparing the classification capacity of a linear discriminant (LD) over the predictions on single trajectory states' labels generated by the t-SNE procedure (left panel) and the HMM approach (right panel). For each given number of states (H values in the case of the t-SNE approach) we show both individual (dashed grey lines) and average (black solid line) results. The larger the F-measure the better match between the LD prediction and the t-SNE or HMM predictions, therefore the more likely the separation into different modes (e.g. the 3-state case). For the HMM approach (right panel), we also show the 1-BIC curve (values scaled normalized from 0 to 1) showing that the larger the number of states one pre-assumes the larger the likelihood of the HMMs but with decreasing increments. In other words, as we keep on adding states the increase in likelihood is smaller.

2. We used the HMM obtained in step 1 to predict the sequence of states of the validation set (the hold out individual).

3. We used the state's labels (the sequence of states obtained in step 1) to fit a LD to the same training data.

4. We used the LD obtained in 3 to predict the state's labels for the validation set (the hold out individual).

5. We compare the state's labels obtained for the hold out data with the two approaches: the HMM (step 2) and the LDA (step 4).

Again, by means of confusion matrices we compared the predictions on the states' labels done by the HMMs and the LDAs for a given number of states and for each individual trajectory. As for the t-SNE analysis, we summarized all the information with the F-measure statistic and, for each number of states, we plotted the individual F-measure values and added an average line (Figure 6, right).

Finally, for the case of the HMMs, we were also able to estimate the log-likelihood, the AIC and the BIC of the models predictions for each individual trajectory. Similarly to what we observed with the full dataset analysis, the individual models obtained from the leave-one-out strategy showed monotonic behaviour (log-likelihood increasing, and AIC and BIC decreasing) with a slight saturation effect as the number of states increased. So again, the log-likelihood values and relatives (AIC, BIC) cannot be used to discriminate between models.

References

1. Bartumeus, F., Raposo, E.P., Viswanathan, G.M. & da Luz, M.G.E. (2014). Stochastic optimal foraging: tuning intensive and extensive dynamics in random searches. *PloS ONE*, 9, e106373.

2. Bazazi, S., Bartumeus, F., Hale, J.J. & Couzin, I.D. (2012). Intermittent motion in desert locusts: Behavioural complexity in simple environments. *PLoS Comput Biol*, 8(5), e1002498. doi:10.1371/journal.pcbi.1002498.

3. Benhamou, S. (2014). Of scales and stationarity in animal movements. *Ecol. Lett.*, 17, 261–272.

4. Bénichou, O., Loverdo, C., Moreau, M. & Voituriez, R. (2011). Intermittent search strategies. *Rev. Mod. Phys.*, 83, 81–129.

5. Berman, G.J., Choi, D.M., Bialek, W. & Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of free moving fruit flies. *J. R. Soc. Interface*, 11, 20140672.

6. Burnham, K.P. & Anderson, D.R. (2002). *Model selection and inference. A practical information-theoretical approach.* Springer-Verlag, Berlin. 2nd ed.

7. Campos, D., Bartumeus, F., Méndez, V. & Espadaler, X. (2014). Reorientation patterns in central-place foraging: internal clocks and klinokinesis. *J. R. Soc. Interface*, 11, 20130859.

8.

Campos, D., Bartumeus, F., Raposo, E. & Méndez, V. (2015). First-passage times in multiscale random walks: The impact of movement scales on search efficiency. *Phys.Rev.E*, 92, 052702.

9.

Campos, D., Méndez, V. & Bartumeus, F. (2012). Optimal intermittence in search strategies under speed-selective target detection. *Phys. Rev. Lett.*, 108, 028102.

10.

Dean, B., Freeman, R., Kirk, H., Leonard, K., Phillips, R.A., Perrins, C.M. & Guilford, T. (2013). Behavioural mapping of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. *J. Roy. Soc. Interface*, 10.

11.

Dusenbery, D.B. (2009). *Living at Micro Scale.* Harvard University Press.

12.

Edwards, A.M. (2011). Overturning conclusions of Lévy flight movement patterns by fishing boats and foraging animals. *Ecology*, 92, 1247–1257.

13.

Edwards, A.M., Phillips, R., Watkins, N.W., Freeman, M.P., Murphy, E.J., Afanasyev, V., Buldyrev, S.V., da Luz, M.G.E., Raposo, E.P., Stanley, H.E. & Viswanathan, G.M. (2007). Revisiting Lévy flight search patterns wandering albatrosses, bumblebees and deer. *Nature*, 449, 1044–1048.

14.

Fronhofer, E.A., Hovestadt, T. & Poethke, H.J. (2013). From random walks to informed movement. *Oikos*, 122, 857–866.

15.

de Jager, M., Bartumeus, F., Kölzsch, A., Weissing, F.J., Hengeveld, G.M.,

Nolet, B.A., Herman, P.M. & van de Koppel, J. (2014). How superdiffusion gets arrested: ecological encounters explain shift from Lévy to Brownian movement. *P. Roy. Soc. London B Bio*, 281 (1774), 2013–2605.

16.
de Jager, M., Weissing, F.J., Herman, P.J., Nolet, B.A. & van de Koppel, J. (2011). Lévy walks evolve through interaction between movement and environmental complexity. *Science*, 332, 1551–1553.

17.
de Jager, M., Weissing, F.J., Herman, P.M.J., Nolet, B.A. & van de Koppel, J. (2012). Response to comment on Lévy walks evolve through interaction between movement and environmental complexity. *Science*, 335, 918.

18.
Jansen, V.A.A., Mashanova, A. & Petrovskii, S. (2012). Comment on Lévy walks evolve through interaction between movement and environmental complexity. *Science*, 335, 918.

19.
Lovely, P.S. & Dahlquist, F.W. (1975). Statistical measures of bacterial motility and chemotaxis. *J. Theor. Biol.*, 50, 477–496.

20.
van der Maaten, L.J.P. & Hinton, G.E. (2008). Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9, 2579–2605.

21.
O'Brien, W.J., Browman, H.I. & Evans, B.I. (1990). Search strategies of foraging animals. *Am. Sci.*, 78, 152–160.

22.
Petrovskii, S., Mashanova, A. & Jansen, V.A.A. (2011). Variation in individual walking behaviour creates the impression of a Lévy flight. *P. Natl. Acad. Sci. USA*, 105, 19052–19059.

23.

Pyke, G.H. (1981). Optimal travel speeds of animals. *Am. Nat.*, 118, 475–487.

24.

Raposo, E.P., Bartumeus, F., da Luz, M.G.E., Ribeiro-Neto, P.J., Souza, T.A. & Viswanathan, G.M. (2011). How landscape heterogeneity frames optimal diffusivity in searching processes. *PLoS Comput Biol*, 7, e1002233.

25.

Reynolds, A.M. (2012). Distinguishing between Lévy walks and strong alternative models. *Ecology*, 93(5), 1228–1233.

26.

Reynolds, A.M. (2014). Mussels realize weierstrassian Lévy walks as composite correlated random walks. *Sci. Rep.*, 4, 4409.

27.

Salvador, L.C.M., Bartumeus, F., Levin, S.A. & Ryu, W.S. (2014). Mechanistic analysis of the search behaviour of *Caenorhabditis elegans*. *J. R. Soc. Interface*, 11(92), 1742–5662.

28.

Schick, R., Loaire, S.R., Colchero, F., Best, B.D., Boustany, A., Conde, D.A., Halpin P. N. Joppa, L.N., McClellan, C.M. & Clark, J.S. (2008). Understanding movement data and movement processes: current and emerging directions. *Ecol. Lett.*, 11, 1338–1350.