SI Appendix: Information use and the evolution of collective migration

Vishwesha Guttal[1] **and Iain D. Couzin**[2]

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, 08544, USA

Contents

A SI Methods: details of model implementation **2**
- A.1 Movement rules: . . . 2
- A.2 Starting conditions: . . . 3
- A.3 Fitness evaluation: . . . 3
- A.4 Selection algorithm: . . . 4
- A.5 Group size calculations . . . 4
- A.6 Parameters and sensitivity: . . . 4

B SI Methods: Evolutionary stable strategy, or optimal strategy, for a solitary individual **5**

C SI Methods: Evolutionary simulations for populations **7**
- C.1 Evolutionary stable strategies (ESS) and evolutionary stable states (ESSt) . . . 7
- C.2 Initial condition or history dependence of ESSt . . . 7
- C.3 Invasibility analysis and 'robust evolutionary stable state' (rESSt) . . . 7
- C.4 Parameter scans . . . 8

D SI Text: The evolution of bimodal strategies and generality with respect to cost function **14**
- D.1 An intuitive explanation of the evolutionary branching process . . . 14
- D.2 Generality with respect to the choice of exponential cost function . . . 15

E SI Figure: Evolutionary outcome as a function of cost of gradient detection and cost of sociality **15**

F SI Text: A model in which individuals can employ their strategy probabilistically **18**

G SI Text: A model in which individuals can employ their strategy in a context dependent way **20**

H SI Text: The microevolutionary response of migration to habitat fragmentation and changes in population density **23**

[1]vguttal@princeton.edu
[2]icouzin@princeton.edu

A SI Methods: details of model implementation

A.1 Movement rules:

Social Interactions: In a population consisting of N individuals, a focal individual i with position vector $\mathbf{c}_i(t)$ and direction vector $\mathbf{v}_i(t)$ moves at a constant speed s at time t. To account for individuals maintaining a personal space, we assume that they avoid collision with their neighbors when they are within a short zone-of-avoidance with radius r_a by moving away from them in the direction:

$$\mathbf{d}_i(t+\Delta t) = -\sum_{j\neq i} \frac{\mathbf{c}_j(t)-\mathbf{c}_i(t)}{|\mathbf{c}_j(t)-\mathbf{c}_i(t)|} \tag{1}$$

where $\mathbf{d}_i(t+\Delta t)$ is the desired direction of travel in the next time step. Avoidance of other individuals is assumed to be of highest priority[1,2]. In the event of no neighbors being detected within the zone-of-avoidance, an individual will be attracted towards and align its direction of travel with other individuals within a local zone-of-socialization r_s:

$$\mathbf{d}_{si}(t+\Delta t) = \sum_{j\neq i} \frac{\mathbf{c}_j(t)-\mathbf{c}_i(t)}{|\mathbf{c}_j(t)-\mathbf{c}_i(t)|} + \sum_{j=1}^{N} \frac{v_j(t)}{|v_j(t)|} \tag{2}$$

where $\mathbf{d}_{si}(t+\Delta t)$ is the desired direction of travel due to social interactions.

Migratory gradient/cue detection: Without loss of generality, we assume the x-axis to be the direction of an environmental gradient. An individual with a gradient detection ability, ω_{gi}, determines the direction of gradient, at each time step, with some errors. The larger the ω_{gi}, the lower the influence of stochastic effects, and thus the higher accuracy at gradient detection. The angular deviation from the migratory direction (x-axis), θ, is determined by $d\theta = -\omega_g\theta dt + \sigma_g dW_g$ where dW_g is the Wiener process. We use the exact expression for the time evolution of this stochastic process, $\theta(t+\Delta t) = \theta(t)e^{-\omega_g\Delta t} + \sqrt{\frac{\sigma_g^2}{2\omega_g}(1-e^{-2\omega_g\Delta t})}\zeta_t$ where ζ_t is the Gaussian white noise. We wrap θ appropriately so that its an angle $\in [-\pi,\pi]$ and obtain migratory direction to be $\mathbf{d}_{gi} = (\cos(\theta)\hat{x} + \sin(\theta)\hat{y})$. We note that the gradient detection process has correlations arising either due to temporal correlations in the gradient, or from inherent errors in an individual's detection ability.

Movement: An individual balances the tendencies of gradient climbing with a desire to be social by weighting them proportionately and moving in the direction[2]:

$$\mathbf{d'}_i(t+\Delta t) = \frac{\omega_{si}\hat{\mathbf{d}}_{si}(t+\Delta t) + \omega_{gi}\hat{\mathbf{d}}_{gi}(t+\Delta t) + \sigma_r\hat{\mathbf{d}}_{ri}(t+\Delta t)}{|\omega_{si}\hat{\mathbf{d}}_{si}(t+\Delta t) + \omega_{gi}\hat{\mathbf{d}}_{gi}(t+\Delta t) + \sigma_r\hat{\mathbf{d}}_{ri}(t+\Delta t)|} \tag{3}$$

where ω_{gi} is the individual gradient detection ability, ω_{si} is the sociality trait, and $\hat{\mathbf{d}}_{ri}$ is a vector with random orientation to simulate inherent errors in individual perception and motion. The hat sign $\hat{}$ on the direction vectors indicate that they are normalized (unit) vectors.

Individuals have a maximum turning rate of θ_{max} and thus can turn at the most $\theta_{max}\Delta t$ radians in the time step Δt. If the angle between their current velocity $\mathbf{v}_i(t)$ and their desired direction $\mathbf{d_i}'(t + \Delta t)$ is less than $\theta_{max}\Delta t$, then the new direction of their movement would be $\mathbf{v}_i(t + \Delta t) = \mathbf{d_i}'(t + \Delta t)$; otherwise, they turn $\theta_{max}\Delta t$ towards it. The new position vector of the individual i is then given by $\mathbf{c}_i(t + \Delta t) = \mathbf{c}_i(t) + s\,\mathbf{v_i}(t + \Delta t)\Delta t$.

Boundary Conditions: We assume that the space is periodic with a length l in each of the two dimensions implying that a particle leaving one side of the simulated environment reappears on the opposite side with the same velocity. Since the space wraps onto itself in each dimension the resulting environment can be mapped onto a torus. Such a boundary condition is a computational technique to simulate a large system, where edge effects are negligible, by focussing on a recurring smaller part of the system that is sufficient to capture the essential biology.

Periodic boundary condition, however, can lead to the following artificial feature when individuals move along a gradient. An individual with a high gradient detection ability, ω_{gi} moves accurately along the positive x-axis with minor stochastic effects in its direction of travel. Owing to periodic nature of the boundary, such individuals return to nearly the same path after crossing the boundary and therefore effectively travel along an one-dimensional recurring path. In order to avoid this potential artefact, all individuals are perturbed along the y-axis at the time they cross the positive x-boundary in space. The strength of perturbation is given by the time-dependent quantity $10 \times s \times cos(2\pi t/t_p)$, where t is the time when the individual crosses boundary. Such a perturbation at the boundary ensures that an individual with high gradient detection ability ω_{gi}, or a group consisting of individuals having strong gradient detection ability, does not follow a straight line path but explores the whole two dimensional space. Yet this preserves the cohesion of a group because any two nearby individuals of a group cross the boundary nearly simultaneously and are therefore displaced by (nearly) same strength of the time-dependent (but not random) perturbation.

A.2 Starting conditions:

In each run, individuals start at random positions in space with random orientations.

A.3 Fitness evaluation:

In each generation, the fitness of individuals is averaged over n_r runs. In any run, from their starting conditions, individuals move in space following above equations of motion for τ_{tr} time steps representing the transient time to form group structures (if any) at the dynamic equilibrium. The fitness is then evaluated for each individual over an interval of τ_{fit} time steps. The fitness consists of a benefit, b_i, defined as the (normalized) distance traveled along positive x-axis (with maximum being 1 when travelling along the direction of gradient and minimum being -1 when travelling opposite to the direction of gradient). The fitness also includes costs incurred by individuals utilizing their gradient detection ability ω_{gi} and sociality ω_{si}: we assign $c_{gi} = p_g(e^{\omega_{gi}/\omega_{gc}} - 1.0)$ and $c_{si} = p_s(e^{\omega_{si}/\omega_{sc}} - 1.0)$. Assuming that the costs and benefits can be measured in the same currency we

define the individual fitness by: $f_i = b_i - c_{gi} - c_{si}$. In the subsequent sections, we refer to p_g as the cost of gradient detection ability and p_s to be the cost of social interactions.

A.4 Selection algorithm:

The two evolvable phenotypes, individual gradient detection ability, ω_{gi}, and sociality, ω_{si}, vary continuously and take nonnegative values. At the beginning of first generation, individuals are assigned phenotypes drawn from a probability distribution (*e.g.,* a distribution in which all have zero gradient detection ability and zero sociality as in Figure 2 of the main text; or chosen from uniformly distributed random numbers in the phenotypic space as in Figure S4; or an evolved state from a different parameter value as in Figure S5). We use Roulette-wheel-selection algorithm [3] where each individual reproduces asexually with a probability that is proportional to its relative fitness value [4]. Offspring carry similar traits to their parents after undergoing a small mutation which is a Gaussian random number with mean zero and standard deviation σ_μ. Generations are nonoverlapping and the number of individuals are constant across generations. The selection process is repeated until a stable distribution of phenotypes is obtained. The stability of evolved state thus obtained is checked by an invasibility analysis (see SI Methods C).

A.5 Group size calculations

We define group as a collection of individuals satisfying the following condition: if two individuals i and j are within their distance of social interaction r_s, then they belong to the same group. Using the equivalence class algorithm [5] we determine all groups and their constituent members at τ_g timesteps and then this is averaged over 10,000 realizations to obtain Figure 2*D* in the main text.

A.6 Parameters and sensitivity:

Unless stated otherwise we have used the following parameter values in our simulations: $l = 2.0$, $dt = 0.2$ of a unit time step (time step), $\theta_{max} = 2.0$ rad per unit time, $\tau_{tr} = 2000$ time steps, $\tau_{fit} = 500$ time steps, $\tau_g = 2500$ timesteps, $\sigma_g^2 = 0.2$, $s = r_a$ per unit time, $n_r = 30$, $\omega_{gc} = 4.0$, $\omega_{sc} = 4.0$, $\sigma_\mu = 0.01$, $r_s/r_a = 6.0$, $\beta = 1.0$ and $\rho = N/(l/BL)^2 = NBL^2/L^2$ where $BL = r_a$ is the body length of the individual. We specify the density, ρ, and the number of individuals N for a simulation and then determine the absolute values of r_s and r_a (also see Table 1).

We have varied many of these parameter values, in particular, τ_{tr}, τ_{fit}, τ_g, n_r, ω_{gc}, ω_{sc}, σ_μ and r_s/r_a. Additionally, we tried other monotonically increasing functions for the cost as the gradient detection ability increases, in particular a linear and a square-root function. These did not affect the qualitative nature of our results (see SI Appendix D.2).

Quantity	Description	Values	Units/Dimensions
l	Size of the continuous space in each dimension	2.0	L
N	Population size	320 or 16,384	Individuals
ρ	Population density	10^{-5} to 1	Individuals per BL^{-2}
r_a	Zone of avoidance or size of a body length (BL)	$l\sqrt{\rho/N}$	L per BL
r_s	Zone of social interactions	$6\,r_a$	L per BL
s	Speed	r_a per unit time	L per $BL\,T$
θ_{max}	Maximum turning angle per unit time	2.0 rad per unit time	T^{-1}
σ_g	Randomness/error in gradient detection	$\sqrt{0.2}$	-
σ_r	Randomness/error in motion	0.1 or 1.0	-
dt	Discrete time step	0.2 of a unit	T
ω_{gi}	Gradient detection ability of individual i	Evolvable	-
ω_{si}	Sociality trait of individual i	Evolvable	-
p_g	(Prefactor of) Cost of gradient detection	0 to 30	F
p_s	(Prefactor of) Cost of sociality	0 to 30	F
ω_{gc}	Scale for cost of gradient detection	4	-
ω_{sc}	Scale for cost of sociality	4	-
c_{gi}	Cost incurred due to gradient detection (ω_{gi})	$c_{gi} = p_g(e^{\omega_{gi}/\omega_{gc}} - 1.0)$	F
c_{si}	Cost incurred due to sociality (ω_{si})	$c_{si} = p_s(e^{\omega_{si}/\omega_{sc}} - 1.0)$	F
β	Degree of habitat fragmentation	See SI Appendix H	-
d_i	Normalized distance traveled along the migratory gradient	See Eq(4)	F
b_i	Benefit of migration	$b_i = d_i^{\beta}$	F
f_i	Fitness of an individual	$f_i = b_i - c_{gi} - c_{si}$	F
τ_{tr}	Transient time from starting conditions	2000 time steps	T
τ_{fit}	Time interval during which fitness is evaluated	500 time steps	T
τ_g	Time period after which group sizes are evaluated	2500 time steps	T
n_r	Number of realizations per generation	30	-
σ_μ	Strength of mutation in the evolvable parameters	0.01	-

Table 1: Summary of model quantities (BL = Body length, L = length, T = time, F = unit of fitness).

B SI Methods: Evolutionary stable strategy, or optimal strategy, for a solitary individual

For an isolated individual the evolved gradient detection ability ω_{gi}^* must optimize the trade-offs between costs of gradient detection (c_{gi}) and resulting benefits of migration (b_i). We first determine the fitness gained by solitary individuals using analytical techniques.

Benefit: First we calculate the migratory benefit an individual gains by having a certain ability to detect the gradient, ω_{gi}. During T units of time the benefit (b_i), defined as the normalized distance traveled along the positive x-axis (d_i), gained by an individual i is:

$$b_i = d_i = \frac{\sum_{t=0}^{T} s\cos(\theta(t))}{T\,r_a} \tag{4}$$

where $\theta(t)$ is the angular deviation of the direction vector from the positive x-axis at time step

t. Since $s = r_a$, the maximum benefit is 1 if an individual travelling along the positive x-axis for all time steps, and the minimum fitness is -1 if the traveling direction is opposite to the gradient (*i.e.,* along the negative x-axis) for all time steps. The desired direction of motion at each time step, for a solitary individual, is given by:

$$\mathbf{d'}_i(t+\Delta t) = \frac{\omega_{gi}\hat{\mathbf{d}}_{gi}(t+\Delta t) + \sigma_r \hat{\mathbf{d}}_{ri}(t+\Delta t)}{|\omega_{gi}\hat{\mathbf{d}}_{gi}(t+\Delta t) + \sigma_r \hat{\mathbf{d}}_{ri}(t+\Delta t)|} \tag{5}$$

where

$$\hat{\mathbf{d}}_{gi}(t+\Delta t) = \cos(\theta(t))\hat{x} + \sin(\theta(t))\hat{y} \tag{6}$$

$$\text{and} \quad \dot{\theta} = -\omega_{gi}\theta(t) + \eta_{gi}(t) \tag{7}$$

with $\eta_{gi}(t)$ being uncorrelated Gaussian noise with mean zero and variance $2\sigma_g^2$ (i.e., $\langle \eta_{gi}(t)\eta_{gi}(t')\rangle = 2\sigma_g^2\delta(t-t')$). The stochastic process described by Eq 7 is the well known Ornstein-Uhlenbeck process and its solution is given by[6,7]:

$$\theta(t) = \theta(0)e^{-\omega_{gi}t} + \int_0^t e^{-\omega_{gi}(t-s)}\eta_{gi}(s)ds \tag{8}$$

for all t.

Assuming that σ_r is small, the average benefit is given by:

$$\langle b_i\rangle = \langle\cos(\theta(t))\rangle \tag{9}$$

$$= \langle\frac{e^{i\theta(t)} + e^{-i\theta(t)}}{2}\rangle \tag{10}$$

$$= \langle e^{i\theta(t)}\rangle \qquad \text{(since } \theta \text{ is a process with mean 0)} \tag{11}$$

$$= \langle e^{i\theta(0)e^{-\omega_{gi}t}}\rangle_{ic}\ \langle e^{i\int_0^t e^{-\omega_{gi}(t-s)}\eta_{gi}(s)ds}\rangle_{\eta_{gi}} \tag{12}$$

$$= \exp[-\frac{\sigma_g^2}{\omega_{gi}}] \tag{13}$$

where the $\langle\rangle$ denotes the averaging over the errors (noise) over many realizations (n_r) of migratory events/simulations.

Cost: As outlined in the SI Methods A, we assume the cost for having a certain gradient detection ability ω_{gi} to be increasing with the ability as follows:

$$c_{gi} = p_g(\exp(\omega_{gi}/\omega_{gc}) - 1.0) \tag{14}$$

Fitness: The average fitness gained by an individual, which is a measure of its reproductive success, is given by:

$$F = \langle F_i\rangle = \langle b_i\rangle - \langle c_{gi}\rangle \tag{15}$$

$$= \exp[-\frac{\sigma_g^2}{\omega_{gi}}] - p_g(\exp[\frac{\omega_{gi}}{\omega_{gc}}] - 1.0) \tag{16}$$

Figure 1 of the main text shows a very good match between average fitness value obtained by numerical simulations and the analytical expression of Eq (16).

C SI Methods: Evolutionary simulations for populations

C.1 Evolutionary stable strategies (ESS) and evolutionary stable states (ESSt)

An 'evolutionarily stable strategy' (ESS) is defined as a strategy such that if all the members in a population adopt it, it can not be invaded by any other mutant strategy [4].

An 'evolutionary stable state' (ESSt) is one that is restored by selection after the introduction of a rare mutant close to the resident population's phenotype. Such a population can be phenotypically, or genetically, monomorphic or polymorphic [4].

In the following subsection we first note that one or more ESSt's may exist under a given parameter conditions in an evolutionary game [4]. In such circumstances we define a 'robust evolutionary stable state' (rESSt) as one that is restored by selection after the introduction of a rare mutant from another ESSt, or more generally, a rare mutant that is far from the resident population's phenotype. Details are discussed below.

C.2 Initial condition or history dependence of ESSt

We observe that the state to which a population evolves may depend on the initial conditions *i.e.,* on the phenotype distribution that the individuals in the population were assigned at the 1^{st} generation. For a low cost of gradient detection ability p_g, as shown in Figure S1, populations evolve to a bimodal state consisting of leaders and social individuals whose characteristics (such as fraction of leaders in the population) quantitatively match very well with each other even though they start from very different initial conditions. On the other hand, when the costs of gradient detection are relatively high we find initial condition dependence: Figure S2 shows that two populations (top two rows: 2(a-d)) quickly moved away from the initial phenotype distribution and eventually underwent an evolutionary branching (see SI Text D). The third and fourth populations (two bottom rows: 2(e-h)), however, evolved to an unimodal state with no leaders but all social individuals. We note that both of these states are stable under small mutations to offspring phenotypes at the beginning of every generation and therefore, are evolutionary stable states (ESSt). More generally we find that there is no initial condition dependence for zero to very low, and for very high, values of costs of gradient detection, p_g. But there is an intermediate region of parameter values p_g where the evolved strategies show a hysteretic or initial condition dependence.

C.3 Invasibility analysis and 'robust evolutionary stable state' (rESSt)

If multiple ESSt's exist for a given parameter condition, as in Figure S2, we are interested in determining which of these multiple ESSt's is more robust. We do so by invading one of the ESSt's thus obtained with rare mutants, including those far from the resident population's phenotype, as described here: we introduce one individual (we have also tried 1% of the population as mutants) having a certain phenotype value (ω_{gI} and ω_{sI}, I standing for an 'Invader') and check if the mutant

population grows in subsequent generations. We perform this check for the entire phenotype space[3]. Additionally, for each of the ESSt's we introduce a small number of mutants from other ESSt obtained for that parameter value. If the ESSt is not invaded by any of these mutant possibilities, we call such an evolved state as a robust evolutionary stable state (rESSt).

Figure S3(a-c) shows that a single mutant with a high gradient detection ability but no sociality, *i.e.,* $\omega_{gI} = 1.0$ and $\omega_{sI} = 0.0$, in a resident population of 16383 social individuals (the evolved state of Figure S2(g-h)), can successfully invade and grow in number. Continuation of these simulations showed that the resulting population eventually equilibrates in a bimodal state consisting of leaders and social individuals. This bimodal state quantitatively matches with the one we obtained in Figure S2(a-d). On the other hand, we find that the bimodal state cannot be invaded by other mutants. In general for our migration model we find that whenever multiple ESSt's are found due to initial condition dependence, then a bimodal state consisting of leaders and social individuals is a robust evolutionary stable state (rESSt); unimodal populations consisting of all leaders or all social individuals are easily invaded by mutants eventually resulting in the bimodal state of leaders and social individuals.

We note that an rESSt may not be an uninvadable state. To determine whether an evolved state is truly uninvadable, one must try invasion with all mutant states including, for example, multimodal states. In the absence of analytical results, such a simulation is not feasible in a system such as ours where individuals can take continuous phenotype values leading to infinitely large combinations of mutant states. Nevertheless the evolved states that we obtain in our simulations are biologically interesting and relevant.

We point out that a relatively faster way to obtain a rEESt is to begin simulations in which individuals are assigned values of ω_{gi} and ω_{si} drawn from a uniformly distributed random variables in the phenotype space[4]. This initial distribution of phenotypes effectively introduces competition between many possible strategies at the very first generation thus potentially enabling the most robust state to evolve. Even though this does not assure that an rESSt will indeed be reached, results in Figure S4 show a bimodal state is evolved in less than 100 generations, in contrast with homogeneous initial condition across populations that could take often more than 1000 generations to reach evolutionary branching (if at all).

C.4 Parameter scans

Clearly, such an invasibility analysis is computationally very expensive and can even be prohibitive if one is interested in a parameter scan as in Figure 3 of the main text. In order to make such

[3]Due to numerical tractability of such an analysis, we often restrict both of these phenotype values in the range $[0, 2]$. Many simulations performed with much higher bounds have all converged to a state within the bounds we have specified. Moreover, results of analytical calculations for solitary individual often provides us with a good reference value for expecting the population level strategies. Therefore, such a restriction on numerical simulations is reasonable and is not expected to change the results we have obtained.

[4]Once again due to numerical tractability of such an analysis, we often restrict both these phenotype values in the range $[0, 2]$

calculations computationally feasible, we perform all of our parameter scan for smaller number of individuals ($N = 320$). We also use the following 'continuation technique' to achieve an ESSt quickly: we begin at a parameter value (for example, given a density, start with the cost $p_g = 0$) and obtain a rESSt through rigorous evaluation through invasibility analysis. To perform the complete parameter scan, we then increment (or decrement) the cost p_g and use the ESSt obtained for the previous cost value as the initial phenotype distribution (*i.e.,* at the 1^{st} generation).

This does not always guarantee a rESSt due to hysteretic effects. We perform a reverse parameter scan by using the ESSt of the previous parameter value as the initial condition for the next one and sample results are shown in Figure S5. Clearly, we see memory effects leading to multiple ESSt's for a given cost of gradient detection. As noted in previous subsection, however, that when multiple ESSt's occur for a given parameter value, a bimodal state is robust (rESSt). Therefore for the parameter scan of density and the cost of gradient detection shown in Figure 3 of the maintext, as well as that of Figure S7, whenever we encountered multiple ESSt's we have determined, and shown, the rESSt among them.

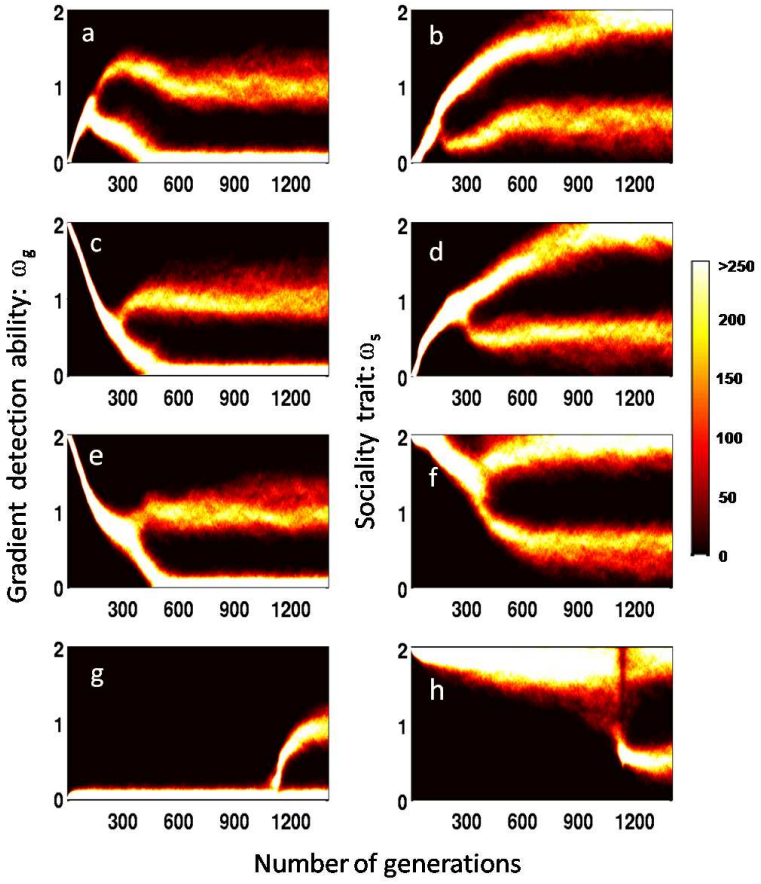


Figure S1: Lack of initial condition (IC) dependence of evolved states at a low cost of gradient detection ability ($p_g = 0.1$). The left column is the temporal evolution of the gradient detection ability, ω_g. The right column shows the sociality trait, ω_s. The colour scale of the plot represents the number of individuals of a given phenotype. Rows represent results of different IC: (a-b) $\omega_{gi} = 0, \omega_{si} = 0$ (c-d) $\omega_{gi} = 2, \omega_{si} = 0$ (e-f) $\omega_{gi} = 2, \omega_{si} = 2$ (g-h) $\omega_{gi} = 0, \omega_{si} = 2$ $\forall$ i at the 1^{st} generation. We find no IC dependence for this set of parameter values (more specifically, for $p_g = 0.1$). Parameter values: $p_s = 0.0, \rho = 2.77 \times 10^{-2} BL^{-2}, \sigma_r = 1.0$ and $N = 16384$.

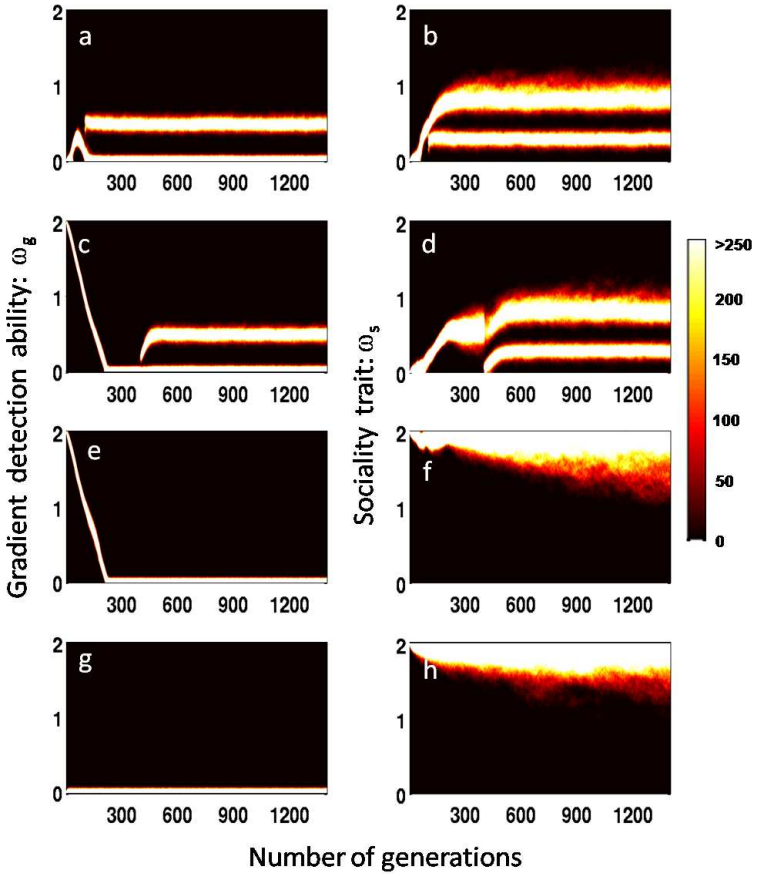


Figure S 2: Initial condition (IC) dependence of evolved states at a moderate cost of gradient detection ability ($p_g = 1.0$). The left column is the temporal evolution of the gradient detection ability, ω_g. The right column shows the sociality trait, ω_s. The colour scale of the plot represents the number of individuals of a given phenotype. Rows represent results of different IC: (a-b) $\omega_{gi} = 0, \omega_{si} = 0$ (c-d) $\omega_{gi} = 2, \omega_{si} = 0$ (e-f) $\omega_{gi} = 2, \omega_{si} = 2$ (g-h) $\omega_{gi} = 0, \omega_{si} = 2 \ \forall \ i$ at the 1^{st} generation. Parameter values: $p_g = 1.0, p_s = 0.0, \rho = 2.77 \times 10^{-2} BL^{-2}, \sigma_r = 1.0$ and $N = 16384$.

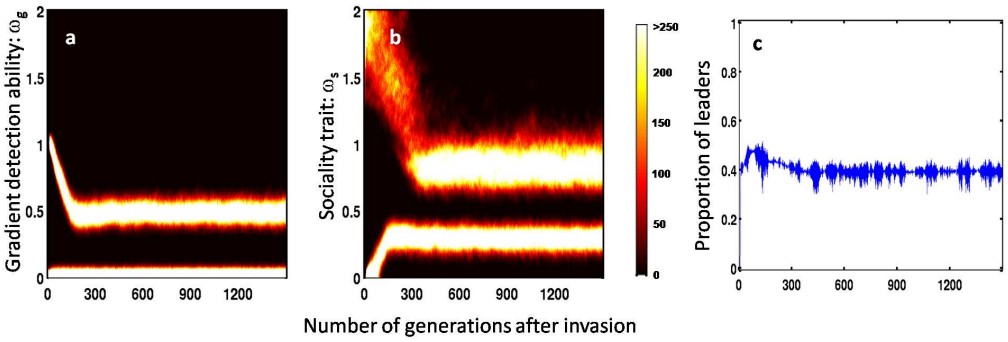


Figure S3: Invasibility analysis. We start from the ESSt of simulation of Figure S2 (g-h) and remove one individual randomly from the population and introduce a leader mutant (high gradient detection ability and low sociality, $\omega_{gI} = 1.0$ and $\omega_{sI} = 0.0$). We then let the evolutionary simulations continue. (a) The evolution of ω_g. (b) The evolution of ω_s (c) State of the population quantified by the fraction of leaders (f_l), defined as those having $\omega_{gi} > 0.5$. At the 1^{st} generation, $f_l = 1/16384 \approx 0.000061$. It rapidly grows and saturates to around 0.4 within a few hundred generations. All parameters are as in Figure S1.

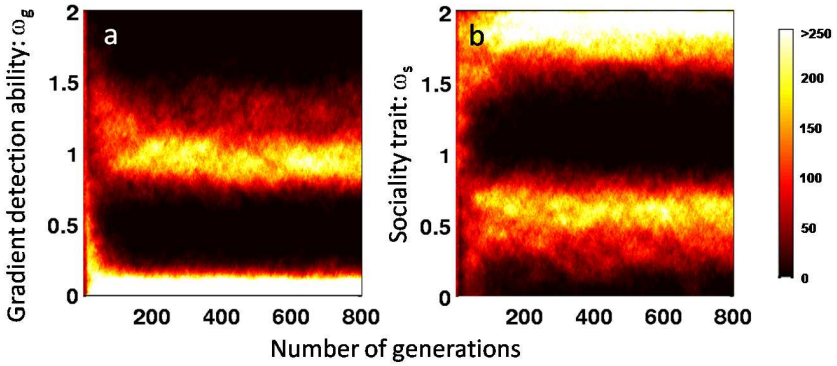


Figure S 4: At the 1^{st} generation, individuals are assigned phenotypes drawn from a uniform random distribution such that $\omega_{gi} \in [0, 2]$ and $\omega_{si} \in [02]$. (a) The evolution of ω_g. (b) The evolution of ω_s. All parameters are as in Figure S1.

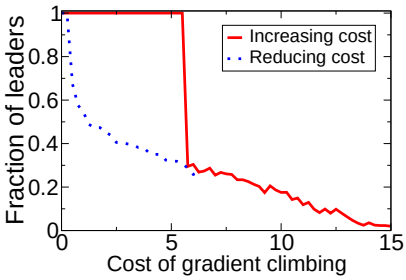


Figure S 5: Hysteresis or memory effects while performing parameter scan. The evolutionary stable state (ESSt) of the population is quantified by the fraction of leaders in the population. The solid red line represents increasing p_g, starting with $p_g = 0$. The dotted blue line is the result for decreasing p_g starting from $p_g = 15.0$. Other parameters are: $p_s = 1.0$, $\rho = 1.77 \times 10^{-3} BL^{-2}$, $\sigma_r = 0.1$, $N = 320$.

D SI Text: The evolution of bimodal strategies and generality with respect to cost function

D.1 An intuitive explanation of the evolutionary branching process

In Figure 2 of the main text we have shown that individuals starting from no gradient detection ability, $\omega_{gi} = 0$ and no sociality, $\omega_{si} = 0$, underwent an evolutionary branching process[8] resulting in a frequency dependent state consisting of leaders (high ω_{gi} and low ω_{si}) and social individuals (non-existing or weak ω_{gi} and high ω_{si}). Here we provide an intuitive explanation for the dynamics leading to this branching process.

In the population with $\omega_{gi} = 0$ and $\omega_{si} = 0 \; \forall \; i$, a mutant with a small and positive value of gradient detection ability, $\omega_{gi} > 0$, will incur a cost but will gain a higher migratory benefits (albeit very small on absolute terms) leading to a net higher relative-fitness than the rest of the population. Therefore the migratory selection pressure acts to increase the gradient detection ability of all individuals. Now if individuals acquire a mutation with a positive sociality trait, $\omega_{si} > 0$, they locally attract each other and others with no-sociality forming small and fragile groups. Yet individuals with a positive gradient detection ability in such groups can detect the migratory direction relatively more accurately due to an averaging process known as the 'many wrongs principle'[9]. Therefore the selection pressure acts to increase the sociality trait, ω_{si} as well. The whole population, therefore, acquires higher values of both the gradient detection ability ω_{gi} and the sociality trait ω_{si} in the initial stages of evolutionary dynamics (Figure 2 of the main text).

At a certain stage in the evolutionary process the social interactions reach a large enough value and begin to influence the evolutionary trajectory in novel ways. In this population individuals with a slightly lower ω_{gi} obtain a relatively higher fitness by incurring lesser costs in the gradient detection process but acquiring similar migratory benefits by social attraction to those with higher ω_{gi}. The decline of gradient detection ability observed in Figure 2 of the main text is a consequence of this reversal of selection pressure on ω_{gi}. Such a process is facilitated by strong social interactions, therefore the increasing trend in sociality trait ω_{si} continues during this stage of evolutionary dynamics.

As the gradient detection ability reduces and sociality trait increases, the population eventually reaches an evolutionary branching point. At this stage mutations that further lower the gradient detection ability and increase the sociality trait continue to be favored due to the same mechanism we explained in previous paragraph. In addition, mutations that increase the gradient detection ability but reduce the sociality trait will also be favored. Such mutants can compensate for the costs incurred in their increased gradient detection ability by the migratory benefits they accumulate. This is facilitated by their reduced tendency to be attracted to other individuals thereby increasing their ability to migrate farther. Eventually the population reaches an evolutionary equilibrium with leaders (those with high ω_{gi} and weak/nonexistent ω_{si}) and social individuals (weak or nonexistent ω_{gi} and high ω_{si}) obtain equivalent fitness values and therefore, coexist with each other in a frequency dependent way.

D.2 Generality with respect to the choice of exponential cost function

It can be easily argued that a bimodal evolutionary stable state would occur even when we chose cost function other than an exponential, as long as it is a monotonically increasing function of ω_g or ω_s. For a moderate value of cost of gradient detection, we will find an optimum gradient detection ability for a solitary individual as seen in Figure 1C of the main text, irrespective of the specific form of the cost function. Consider a homogeneous population in which all individuals have that optimum gradient detection ability. In this population, a social individual mutant with a low ω_g (say, $\omega_g = 0$) but a high sociality trait (ω_s) would have higher fitness because it acquires benefits equivalent to the average of the rest of the population of but its investment in gradient detection is lower. Therefore, its frequency in the population will increase. At the other extreme, in a homogeneous population of $\omega_g = 0$ and high ω_s, a leader mutant with high ω_g but low ω_s is able to obtain migratory benefits that exceeds the investment in gradient detection and hence acquires higher fitness than the resident population. Therefore, the leader mutant will increase its frequency in the population. The above argument for the frequency dependent evolutionary dynamics holds irrespective of the specific form of the cost function and therefore we expect that a coexisting strategy of leaders and social individuals will emerge even when we chose monotonically increasing cost functions other than an exponential.

As a test, we performed sample simulations with a linear cost function ($c_{gi}(\omega_{gi}) = p_g \times \omega_{gi}/\omega_{gc}$) as as well a decelerating square-root function ($c_{gi}(\omega_{gi}) = p_g \times \sqrt{\omega_{gi}/\omega_{gc}}$). Figure S6 shows that these did not alter the qualitative features of our results (that three migratory states of individual migration, collective migration and resident population continue to occur depending on relative costs and benefits).

E SI Figure: Evolutionary outcome as a function of cost of gradient detection and cost of sociality

Social interactions can occur at a cost too, and we include this by assuming that the cost incurred by an individual (c_{si}) increases monotonically with the strength of the sociality trait, ω_{si}; more specifically, $c_{si} = p_s(\exp(\omega_{si}/\omega_{sc}) - 1.0)$. For a given value of p_g and p_s, we quantify the evolved state (rESSt) by the proportion of leaders, defined as those in the high gradient detection ability mode/branch, in the population. We also evaluate the migratory ability of the population by averaging the benefits accumulated by all individuals.

As in Figure 3 of main text we find three qualitatively different migratory states. When the cost of gradient detection is very high no individual evolves to use gradient information and thus there is no migration (Figure S7a-b; dark regions). These resident populations may evolve sociality through neutral mutations (*i.e.,* those which neither increase nor decrease individual fitness) in the evolutionary process leading to non-migrating swarms when the cost of sociality is zero or negligible; or they will consist entirely of asocial individuals leading to solitary random walking individuals even if p_s relatively small.

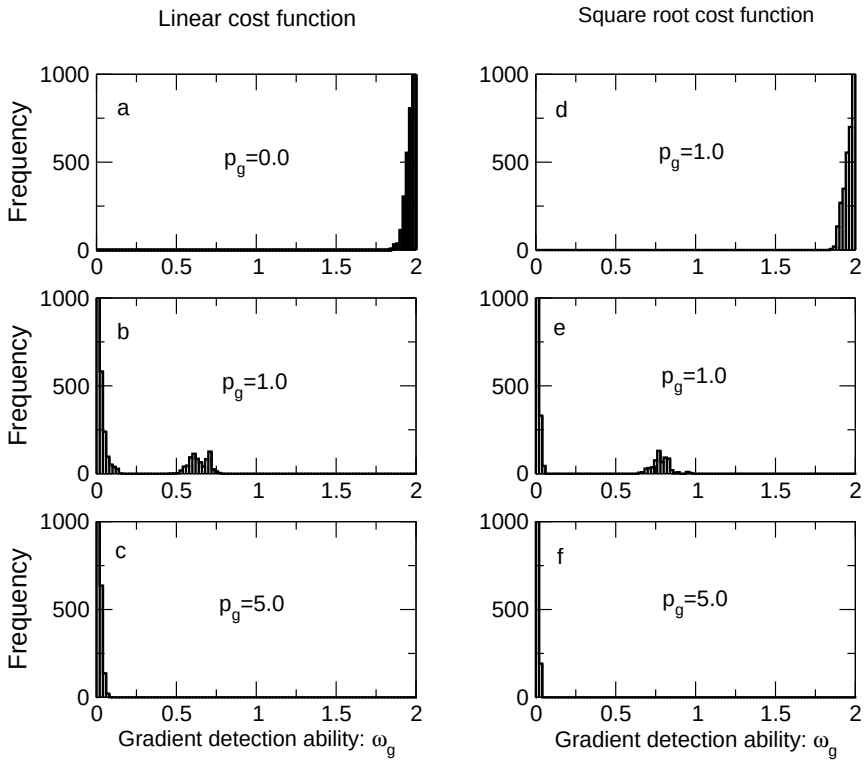


Figure S6: Evolutionary stable states (rESSt) obtained at the end of 300 generations, starting from a random and uniform distribution of phenotypes ($\omega_{gi} \in [0, 2]$ and $\omega_{si} \in [02]$), for different cost functions. In the first column (a-c), we have used linear cost functions: $c_{gi} = p_g\omega_{gi}/\omega_{gc}$ and $c_{si} = p_s\omega_{si}/\omega_{sc}$. In the second column (d-f), we have used a decelerating square root cost function: $c_{gi} = p_g\sqrt{\omega_{gi}/\omega_{gc}}$ and $c_{si} = p_s\sqrt{\omega_{si}/\omega_{sc}}$. Parameter values: $\rho = 2.7 \times 10^{-2}BL^{-2}, \sigma_r = 1.0, N = 320, p_s = 0$. For brevity we have omitted showing the evolved distributions of sociality trait, ω_s.

At the other end, when the cost of gradient detection is very low all individuals evolve to be leaders (Figure S7a; bright region). Individuals in these populations will also evolve sociality thus resulting in collective migration only when the costs associated with sociality is negligible or very low. For low to very large costs of sociality, however, leaders are asocial leading to solitary migration. In a relatively large intermediate region of p_g and p_s, however, leaders (who are typically asocial) and social individuals coexist and thus populations exhibit collective migration.

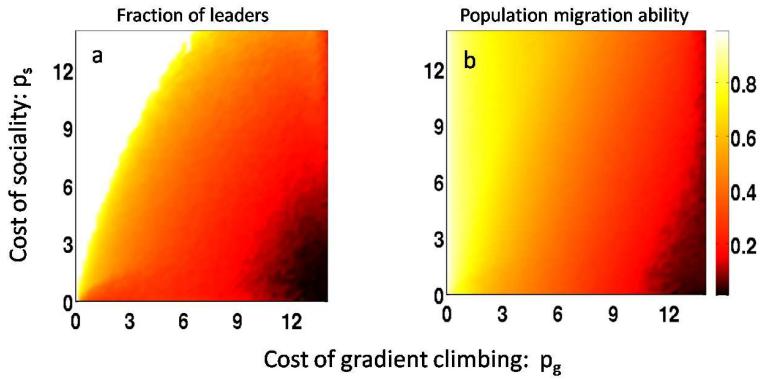


Figure S7: Robust evolutionary stable states (rESSt) as a function of the gradient detection (p_g) and social costs (p_s). (a) The colour scale of the plot represents the rESSt as quantified by the proportion of leaders in the population. (b) The colour scale of the plot represents the migratory ability of the evolved population. Parameter values: $\rho = 7.0 \times 10^{-3} BL^{-2}, \sigma_r = 0.1, N = 320$.

F SI Text: A model in which individuals can employ their strategy probabilistically

So far we have assumed that individuals with a certain strategy use it all times within their lifetime/generation. In nature, however, organisms may employ strategies in more complex ways; for example, an individual may occasionally switch off the ability to perform gradient detection and exploit others' detection by following social cues. We incorporate this feature in the following way.

First we assume that an individual may possess a gradient detection ability denoted by ω_{gi} and the individual chooses to employ it ('switch on') with a probability f_{gi}. Consequently, the individual will switch off its detection ability with a probability of $1 - f_{gi}$. For each individual in the population, the gradient detection 'switch' is updated probabilistically after every τ_f time steps. The equation of motion in this modified model is given by:

$$\mathbf{d'}_i(t + \Delta t) = \frac{\hat{\mathbf{d}}_{si}(t + \Delta t) + \omega_{git}\hat{\mathbf{d}}_{gi}(t + \Delta t) + \sigma_r\hat{\mathbf{d}}_{ri}(t + \Delta t)}{\hat{\mathbf{d}}_{si}(t + \Delta t) + \omega_{git}\hat{\mathbf{d}}_{gi}(t + \Delta t) + \sigma_r\hat{\mathbf{d}}_{ri}(t + \Delta t)|} \tag{17}$$

where the index t in ω_{git} denotes the current status of the gradient detection ability employed by the individual and rest of the symbols are as in the main model used in this paper (see SI Methods A and C). In order for the evolutionary analysis to be tractable, we are measuring the gradient detection ability of individuals relative to their sociality trait by setting $\omega_{gi} = 1.0$ for all individuals i. The relevant phenotypic space for this model is (ω_{gi}, f_{gi}) where ω_{gi} is a continuous variable that can take any nonnegative value and f_{gi}, being a probability, is restricted to the interval $[0, 1]$. Individuals pay a cost that increases exponentially as a function of the gradient detection ability they possess, but only during the intervals of time when they are using that strategy.

For simplicity, we assume that there is no cost associated with switching between strategies. Including this detail could amount to rescaling our existing cost structure and hence based on our results of parameters in previous sections (Figure 3 of the main text and Figure S7) we do not expect the qualitative nature of these conclusions to be affected by such a simplification. Moreover adding a switching cost would provide a strong disincentive (evolutionarily, that is) for the organisms to switch between strategies and hence making the evolution of bimodal/structured populations more likely.

Results of robust evolutionary stable states obtained through numerical simulations are shown in Figure S8(a-c). As in the simpler model presented for the main text, for the intermediate costs of gradient detection the population evolves to a frequency-dependent bimodal state: in one mode the individuals never, or very rarely, use their gradient detection strategy ($f_{gi} \approx 0$) and therefore the precise value of gradient detection ability they have is irrelevant. The other mode consists of individuals employing their gradient detection ability, ω_{gi}, nearly always ($f_{gi} \approx 1$) and they have a finite ω_{gi}. In other words, the population effectively has a certain proportion of population who nearly always employ their finite ability to detect the gradients, referred to as leaders, and rest of the population employs no gradient detection strategy but a socially facilitated movement behavior.

We quantify the measurable impact of two phenotypes used in this model by defining a reduced

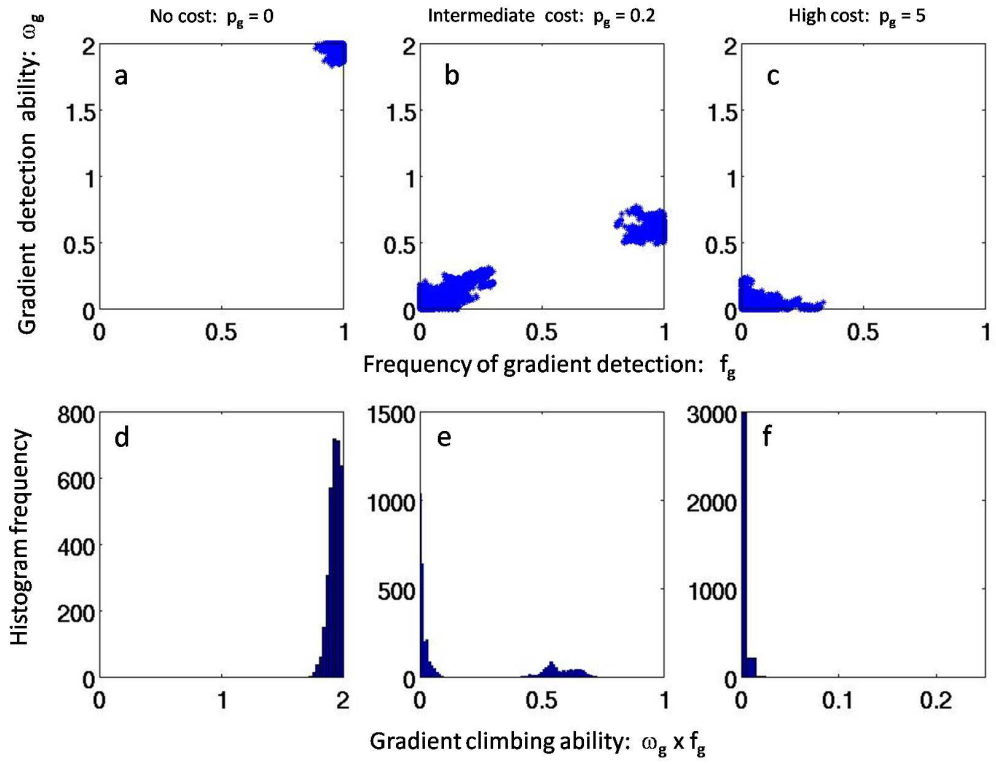


Figure S8: The evolution of migratory strategies in the probabilistic gradient detection model. The top row (a-c): evolved states at the 1250^{th} generation starting from an uniform random distribution of phenotypes (*i.e.*, $\omega_{gi} \in [0, 4]$ and $f_{gi} \in [0, 1]$). (a) With no cost of gradient detection: $p_g = 0.0$. (b) Intermediate cost of gradient detection: $p_g = 0.2$. (c) High cost of gradient detection: $p_g = 5.0$. The bottom row (d-f): the evolved states for the same cost parameter values in terms of a reduced phenotypic space that can be interpreted as the gradient climbing ability. Other parameter values: $N = 320, \rho = 6.67 \times 10^{-3} BL^{-2}, \sigma_r = 0.1, \tau_f = 5\Delta t$

phenotype that is a product of the gradient detection ability, ω_{gi} and the frequency of its usage, f_{gi}, which effectively determines the environmental gradient climbing ability of individuals (hence referred to as 'the gradient *climbing* ability'). These are plotted in Figure S8(d-f) and show a bimodal evolutionary stable state for the intermediate cost value. These results are qualitatively similar to the main results we have presented demonstrating that including more complex rules where individuals can employ their strategies in a probabilistic way does not alter the main conclusions of our paper.

G SI Text: A model in which individuals can employ their strategy in a context dependent way

It is also reasonable, and important, to consider a scenario where individuals might be able to modify their interaction rules depending on the local conditions they are experiencing [10]. For instance, in order to avoid being exploited by social individuals, and/or to exploit others with gradient detection ability, an individual may not perform gradient detection when the local condition is crowded despite possessing a very high gradient detection ability (ω_g). Additionally, this approach allows us to In this section we show that including the possibility for individuals to evolve such context-dependent strategy does not affect the qualitative nature of the results we have obtained.

In this context dependent model an individual i at time t has an ability to detect the gradient, ω_{git}, and will employ it only if the number of nearby neighbours, defined as those within the zone of socialization, r_s, are less than a certain threshold, denoted by n_{gi}, at that time. In other words n_{gi} represents the 'switch-off threshold' for the gradient detection ability, for example based on a quorum detection mechanism [11]. Individuals update such a (local) context dependent ω_{git} every τ_f time steps and follow Eq (17) for their motion. Therefore each individual has two evolvable phenotypes, ω_{gi} and n_{gi}, where ω_{gi} is a continuous variable that can take any nonnegative value and n_{gi} can take any nonnegative integer values[5]. As in SI Text F we assume that: (i) We are measuring the gradient detection ability relative to sociality by setting $\omega_{si} = 1$ for all individuals i. (ii) Individuals pay a cost that exponentially increases as a function of the gradient detection ability they possess, but only during the intervals of time when they are using that strategy. (iii) There is no cost associated with switching between strategies or to measure the local crowding conditions. Note that our argument from SI Text F - that including switching cost does not affect the qualitative nature of our results - continues to hold.

Figure S9(a-c) shows the results of evolutionary simulations of this model. When there is no cost associated with the gradient detection all individuals evolve to a high gradient detection ability, ω_{gi} and a high switch-off/quorum threshold (*i.e.,* number of neighbors) n_{gi}. Since the evolved n_{gi} is higher than the maximum possible number of individuals within the zone of socialization, individuals in this evolved populations always employ their gradient detection ability, ω_{gi} and thus, travel in the direction detected by them after balancing it with their social tendencies. At the other extreme when the cost of gradient detection is very high, no individual evolves to have any significant gradient detection ability thus leading to no migratory behavior. For intermediate values of cost (Figure S9(b)), however, the population evolves to a bimodal state with one mode containing individuals having, no or a very small, n_{gi} and therefore they barely use their ω_g (social individuals). The other mode contains individuals with a very high n_{gi} and therefore they always employ their ω_{gi} (leaders).

We note that the phenotypic space used in this model is relatively complex: a high ω_{gi} or a high n_{gi} alone will not lead to a better gradient climbing or migratory ability. It is a combination

[5] We restrict $\omega_{gi} \in [0, 4]$ and $n_{gi} \in [0, 100]$ for computational tractability

of these two phenotypes that results in the migratory or gradient climbing ability of individuals. Therefore we focus on a behaviorally measurable quantity by reducing the phenotypic space into a more intuitive one dimensional space of the 'gradient climbing ability' as in the previous SI Text F. We make an explicit calculation of how frequently the evolved individuals employ their ω_{gi} by continuously tracking its state in our simulations. We denote the value of frequency thus obtained by f_{gi}. We multiply this quantity with the gradient detection ability, ω_{gi}, to obtain the gradient climbing ability. As we vary the cost parameter, the evolved structure of this more intuitive phenotype shows features that are qualitatively similar to results presented using a much simpler model in the main text of the paper.

We provide an intuitive explanation for the mechanism that maintains the evolutionary stability of bimodal state even when switching between strategies is allowed. We compare two strategies, both having a high ω_{gi} but they differ in their n_{gi}; one with a moderate value of quorum threshold n_m and hence uses the detection strategy when there are not too many individuals around, and the other with a high quorum threshold n_h and hence uses it almost all the time. We now consider the effectiveness of these two individual strategies when they encounter a group in which no other individual has a gradient detection ability (*i.e.,* $\omega_{gi} = 0 \; \forall \; i \in$ group). If the group size is smaller than n_m, both individuals employ their ω_{gi}, emerge at the front due to self-sorting process[1,2] and eventually may move out of the group thus accumulating migratory benefits. If the group size is larger than n_m, however, the individual with moderate value of quorum threshold may never switch on its gradient detection. In contrast, the individual with a higher quorum threshold is likely split from the group and move along the environmental gradient thus accumulating better fitness.

The above argument holds even if the encountered group consisted of individuals all with a high ω_{gi} and a moderate quorum threshold n_m; because if the groupsize is larger than n_m, all individuals are likely to switch of their gradient detection and hence there is no migration. In this group, however, the individual with a high n_{gi} will split from the group and migrate therefore gaining better fitness. In the extreme scenario when all members of a group have high ω_{gi} and a high quorum threshold, an individual with very low n_{gi} can exploit others in the group better than an individual with a moderate n_{gi}. Therefore individuals who employ their gradient detection in a context dependent way (*i.e.,* moderate values of n_{gi}) are outperformed by both the strategies of 'leaders' (*i.e.,* those who always employ their gradient detection) and/or 'naive' individuals (*i.e.,* those who never employ gradient detection). We note that the explicit spatiotemporal dynamic plays a key role in maintaining the evolutionary stability of frequency-dependent bimodal states.

The convergence of qualitative results in models starting from very simple to more complex representation of the world, as well as wide range of parameter scans involving the costs of gradient detection and the costs of sociality, and the benefit structures, shows the generality and potential wide applicability of the central conclusions of this paper.

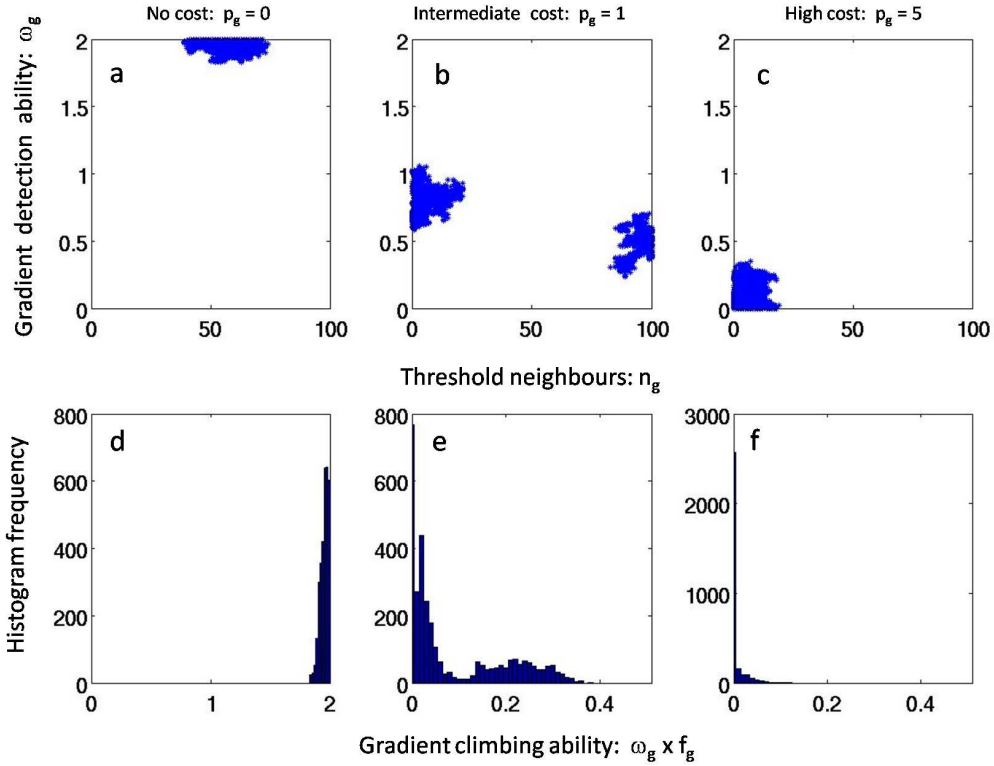


Figure S 9: The evolution of migratory strategies in the context dependent gradient detection model. The top row (a-c): evolved states at the 1000^{th} generation starting from an uniform random distribution of phenotypes (*i.e.*, $\omega_{gi} \in [0, 4]$ and $n_{gi} \in [0, 100]$). (a) With no cost of gradient detection: $p_g = 0.0$. (b) Intermediate cost of gradient detection: $p_g = 1.0$. (c) High cost of gradient detection: $p_g = 5.0$. The bottom row (d-f): the evolved states for the same cost parameter values in terms of a reduced phenotypic space that can be interpreted as the gradient climbing ability. Other parameter values: $N = 320, \rho = 6.67 \times 10^{-3} BL^{-2}, \sigma_r = 0.1, \tau_f = 5\Delta t$

H SI Text: The microevolutionary response of migration to habitat fragmentation and changes in population density

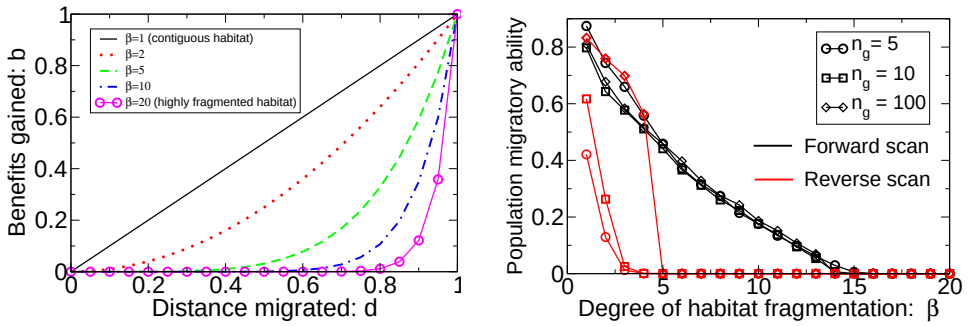
In this section we provide a detailed description of calculations pertaining to the microevolutionary response of migratory phenomenon to habitat fragmentation (Figure 4 in the main text) and changing population densities. We first define how we implement the cost-benefit structure for the gradient detection and migration in a fragmenting habitat. This is followed by how we interpret the microevolution within our model. We then present results across a parameter scan corresponding to habitat fragmentation and changing density.

Migratory benefits and costs: So far we have (implicitly) considered a contiguous habitat where the benefits of migration are assumed to be proportional to the distance travelled along the migratory route (*i.e.,* positive x-axis in our simulations). However, we also want to consider more complex environments such as those that are discontinuous, or fragmented. If the environment becomes increasingly fragmented, individuals may encounter stop-over and/or refueling sites proportionately less frequently and thus they need to cover disproportionately larger distances before accumulating benefits[12,13]. We implement this in the following simple way in our model: the benefit gained by an individual i, b_i, is a nonlinear function of the average (normalized) distance migrated d_i ($0 \leq d_i \leq 1$: see Eq 4 of SI Methods B). More specifically, we assume $b_i = d_i^{\beta}$ where the nonlinearity index β is treated as a proxy for the degree of habitat fragmentation. If $\beta = 1$, we have a contiguous habitat and therefore individuals acquire benefits that are proportional to the distance covered along the migratory direction. As shown in Figure S10(a) the larger the value of β, the larger the nonlinearity, and hence, the organism must cover longer migratory distance to gain benefits. We assume the same cost structure for the gradient detection ability ω_{gi} and sociality ω_{si} as was done for previous calculations: $c_{gi} = p_{gi}(\exp(\omega_{gi}/4.0) - 1.0)$ and $c_{si} = p_{si}(\exp(\omega_{si}/4.0) - 1.0)$.

The microevolutionary response: Empirical evidence suggests that migratory phenomenon can exhibit rapid microevolutionary changes on relatively short ecological time scales such as decades[14,15,13,16]. Human induced ecological changes such as alterations in habitat structure and the density of populations are also likely to occur on these time scales. Our interest is in predicting the 'microevolutionary' response of migratory phenomenon to such changing conditions using our model framework.

To do so, we emphasize that the ESSt's obtained in evolutionary simulations can be history, or initial condition, dependent (Figures S2 and S5). Furthermore, we have shown that although multiple ESSt's may exist for a given set of parameter values, one of those population strategies is likely to be more robust (rESSt); for example, a single but *large mutant* may often change the evolutionary trajectory of the population (see Figure S3 of SI Methods C). In simulating the microevolutionary response of populations we note that *large mutants won't occur within relatively short ecological time scales* over which habitat changes are likely to occur. Therefore, the hysteresis, which we overcame through large mutants while obtaining uninvadable ESSt for the parameter scans (Figure 3 of the main text and Figure S7), is now the key feature[4].

More specifically we begin the population at a given density in a contiguous habitat, ($\beta = 1$)



(a) (b)

Figure S10: (a) Benefits as a function of distance migrated for different degrees of habitat fragmentation (β). (b) Variation in hysteresis curve of the microevolutionary response for different n_g, the length of generations available for the individuals to adapt their strategies, for a given change in ecological parameter. The symbols connected by black lines represent the increasing habitat fragmentation where as the ones connected by red lines correspond to habitat restoration. Different symbols, on the other hand, correspond to different values of n_g for which the simulations were run. $N = 320, \rho = 9.0 \times 10^{-4} BL^{-2}, \sigma_r = 0.10, p_g = 1.0, p_s = 1.0$ and the strength of mutation $\sigma_\mu = 0.01$.

and obtain the rESSt. We then introduce ecological changes by making a small increment in β. As in previous parameter scans (see SI Methods C for methods), we use the evolved state from the previous parameter value as the initial phenotype distribution for this new parameter value and we let the population undergo evolutionary dynamics for n_g number of generations. We continue these simulations until the habitat fragmentation reaches a sufficiently large value where the migration collapses. We then reverse the ecological conditions by restoring the habitat and determine the evolutionary response of the migratory strategies: this is done by gradually decreasing the value of β and as before, we use the evolved phenotype distributions at the previous value of β as the initial distribution for the new one so that we can measure whether we can recover the migratory phenomena. This procedure was followed to obtain Figure 4 of the main text that led to our prediction that the decline of migration will often be relatively gradual in response to habitat fragmentation, but that it will require significantly greater restoration of habitats to recover lost migratory behavior.

We interpret n_g, the number of generations available for the individuals to evolve the strategies for a given change in ecological conditions, as a relative measure of evolutionary time scales to ecological time scales. If $n_g \to \infty$, then the ecological conditions vary extremely slowly, and hence, one can argue that the population reaches the rESSt. As $n_g \to 1$, the evolutionary time scales are comparable to ecological time scales and any resulting adaptations by individuals in that relatively smaller time scale can be treated as microevolution. For the simulations in Figure 4 of main text, we have used $n_g = 300$ generations. Here, we compare how the hysteresis curve changes as we

change n_g. From Figure S10(b) we find that the strong hysteresis effect we reported continues to hold for $n_g = 100, 10$ and 5 generations. In fact, we find that as the evolutionary time scales become comparable to ecological time scales, the hysteresis is more pronounced.

We have carried out full parameter scans of habitat fragmentation and variations in density of populations to determine the microevolutionary response of the migratory phenomena. Results are shown in Figure S12. As noted in the main text, population migratory ability shows strong hysteresis requiring much larger restoration of habitat than at which it first declined. This is due to lack of mutations that exceed a threshold on ecological timescales. We have quantified this threshold as follows: For a given value of habitat fragmentation (β) all individuals in the population, except one, are assigned $\omega_{si} = 0$ and $\omega_{gi} = 0$. We assume that the remaining one individual has a different strategy given by $\omega_{si} = 0$ and $\omega_{gi} = \omega_{gt} > 0$ (for example, due to a mutation). Through simulations we check whether the mutant will grow in the population, or not. We define threshold mutation as the minimum value of ω_{gt} such that the number of mutants in the population has increases (*i.e.,* greater than one) at the end of 10 generations. The threshold ω_g thus determined is plotted Figure S11. Clearly, this threshold value of mutation rises rapidly as with increasing habitat fragmentation. Such large mutations, however, can not occur on ecological timescales of habitat restoration. Therefore we find hysteresis behavior. We also find hysteresis curves as a function of density of population, although to a lesser degree, as seen in Figure S12.

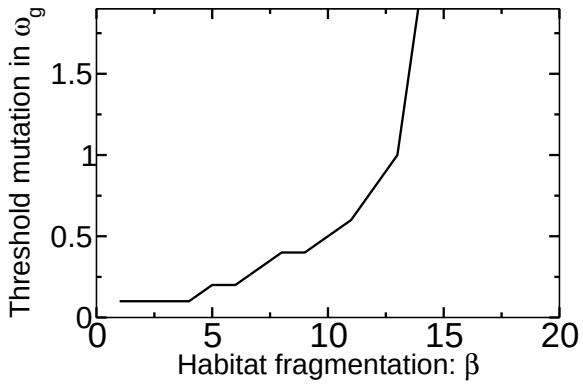


Figure S11: The threshold mutation in ω_g as a function of habitat fragmentation. Parameters: $p_g = 1.0, p_s = 1.0, \sigma_r = 0.1, N = 320$.

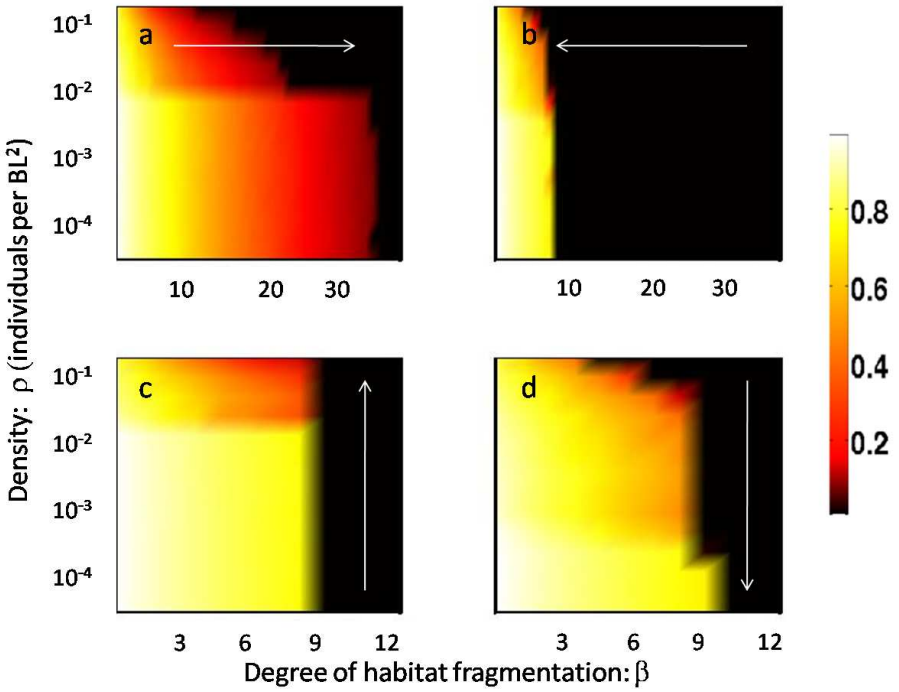


Figure S12: Response of migration to anthropogenic activities. In all of the above plots, the migratory performance of evolved populations is plotted as a heat-map, with brighter regions representing effective migration and dark regions representing no or weak migration (see the color-scale). The x-axis is β, the degree of habitat fragmentation. The arrows indicate the direction in which the parameters were changed. (a) shows how the migratory behavior is affected, for a fixed density, as the habitat fragmentation increases whereas in (b) the habitat is restored. (c) Increasing, or (d) decreasing, density for fixed levels of fragmentation. Parameters: $p_g = 1.0, p_s = 1.0, \sigma_r = 0.1, N = 320$.

References

1. Couzin, I., Krause, J., James, R., Ruxton, G. & Franks, N. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.* **218**, 1–11 (2002).

2. Couzin, I., Krause, J., Franks, N. & Levin, S. Effective leadership and decision-making in animal groups on the move. *Nature* **433**, 513–516 (2005).

3. Goldberg, D. & Deb, K. A comparative analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms* **1**, 69–93 (1991).

4. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, UK, 1982).

5. Press, W., Teukolsky, S., Vetterling, W. & Flannery, B. *Numerical recipes in C* (Cambridge Univ. Press, Cambridge, MA, USA, 1992).

6. Gardiner, C. W. *Handbook of stochastic methods for Physics, Chemistry and the Natural Sciences.* (Springler-Verlag, 2003), 3rd edition edn.

7. Van Kampen, N. *Stochastic processes in physics and chemistry* (North-Holland, 2007).

8. Geritz, S., Kisdi, E., Mesze´ NA, G. & Metz, J. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* **12**, 35–57 (1998).

9. Simons, A. Many wrongs: the advantage of group navigation. *Trends Ecol. Evol.* **19**, 453–455 (2004).

10. Torney, C., Neufeld, Z. & Couzin, I. D. Context-dependent interaction leads to emergent search behavior in social aggregates. *Proc. Nat. Acad. Sci., USA* **106**, 22055–22060 (2009).

11. Surette, M., Miller, M. & Bassler, B. Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: a new family of genes responsible for autoinducer production. *Proc. Nat. Acad. Sci., USA* **96**, 1639–1644 (1999).

12. Houston, A. Models of optimal avian migration: state, time and predation. *J. Avian Biol.* **29**, 395–404 (1998).

13. Alerstam, T., Hedenstrom, A. & Akesson, S. Long-distance migration: evolution and determinants. *Oikos* **103**, 247–260 (2003).

14. Holt, R. The microevolutionary consequences of climate change. *Trends Eco. Evol.* **5**, 311–315 (1990).

15. Berthold, P., Helbig, A., Mohr, G. & Querner, U. Rapid microevolution of migratory behaviour in a wild bird species. *Nature* **360**, 668–669 (1992).

16. Pulido, F. & Berthold, P. Microevolutionary response to climatic change. *Adv. Ecol. Res.* **35**, 151–183 (2004).

Acknowledgements: We thank Simon Levin for useful discussions and comments on the stability of evolutionary states. We are grateful to Simon Garnier for the help in producing videos.