

## **Supplementary Information**

for

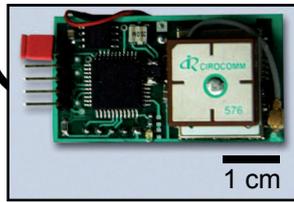
*Máté Nagy, Zsuzsa Ákos, Dora Biro & Tamás Vicsek:*  
Hierarchical group dynamics in pigeon flocks

# 1. Supplementary Figures and Legends

## 1. Free and homing flights of pigeon flocks



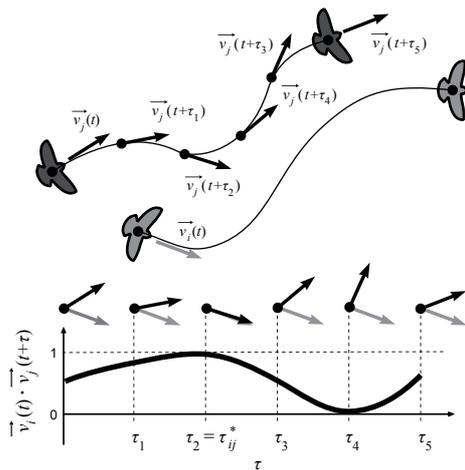
recorded with high-resolution miniature GPS device



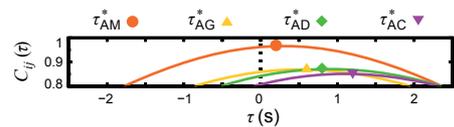
## 2. Reconstruction of 3D flight trajectories



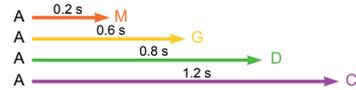
## 3. Calculation of correlation functions in each pair of birds' movement directions in the horizontal plane



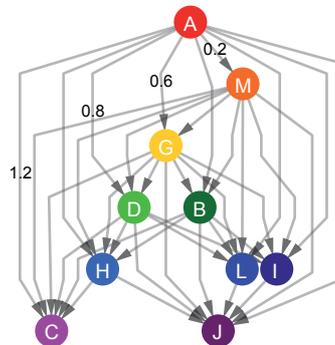
## 4. Allocation of the maximum value of the correlation function as time delay



## 5. Use of time delays in assigning directed leader-follower links



## 6. Construction and analysis of directed hierarchical networks



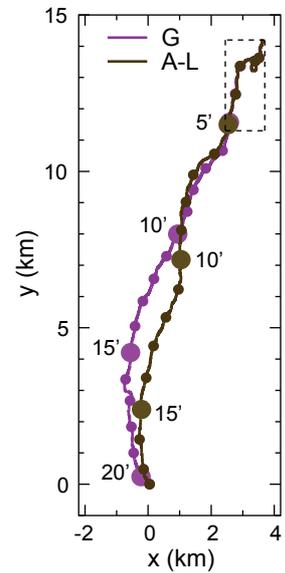
### Supplementary Figure 1

Schematic illustration of the separate stages of the study, from the acquisition of flight trajectory data in homing pigeon flocks, to the analysis of hierarchical leadership networks within such groups.

**a**

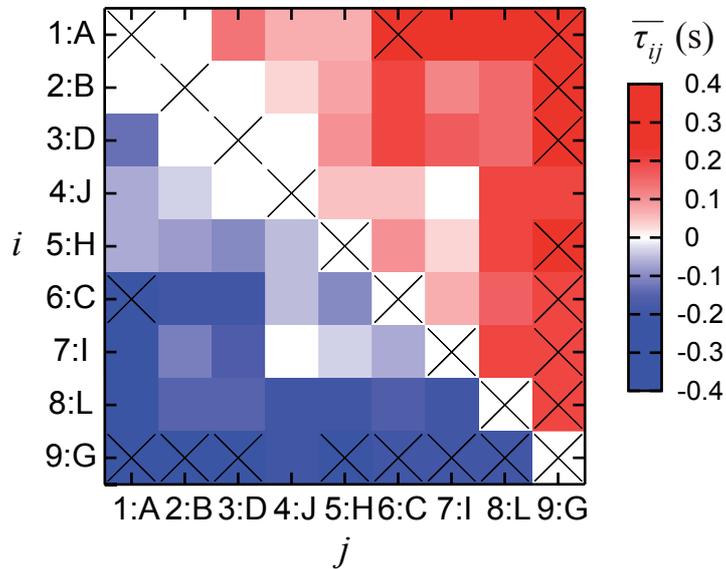


**b**



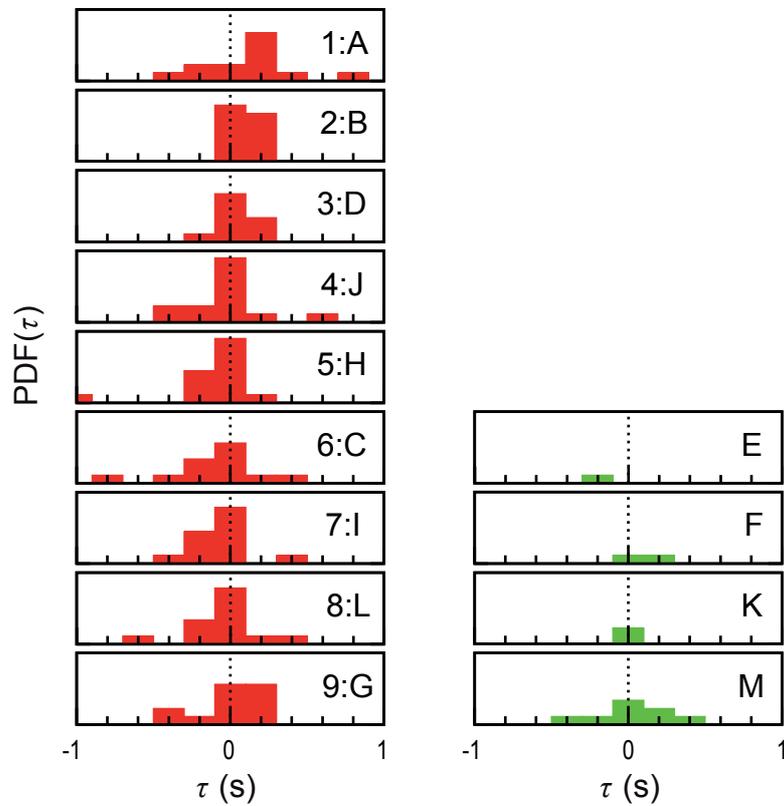
### Supplementary Figure 2

**a**, Subject wearing the custom-made elastic harness used to equip pigeons with the GPS device during all recorded free and homing flights. The device is enclosed within a cloth backpack attached to the harness. **b**, The full length of a group homing flight performed by a flock of nine pigeons. The same flight is animated in Supplementary Movie 2; the region enclosed by the dotted rectangle contains the animated segment, and corresponds to the inset in Supplementary Movie 2. Eight birds (A to L) are shown in brown, the ninth bird (G) is shown in purple. Bird G split from the group approximately a quarter of the way through the journey and returned home on a different trajectory from the rest of the group. The smaller and the larger dots indicate every 1 and 5 minutes of flight time, respectively. The home loft is located at coordinates 0,0.



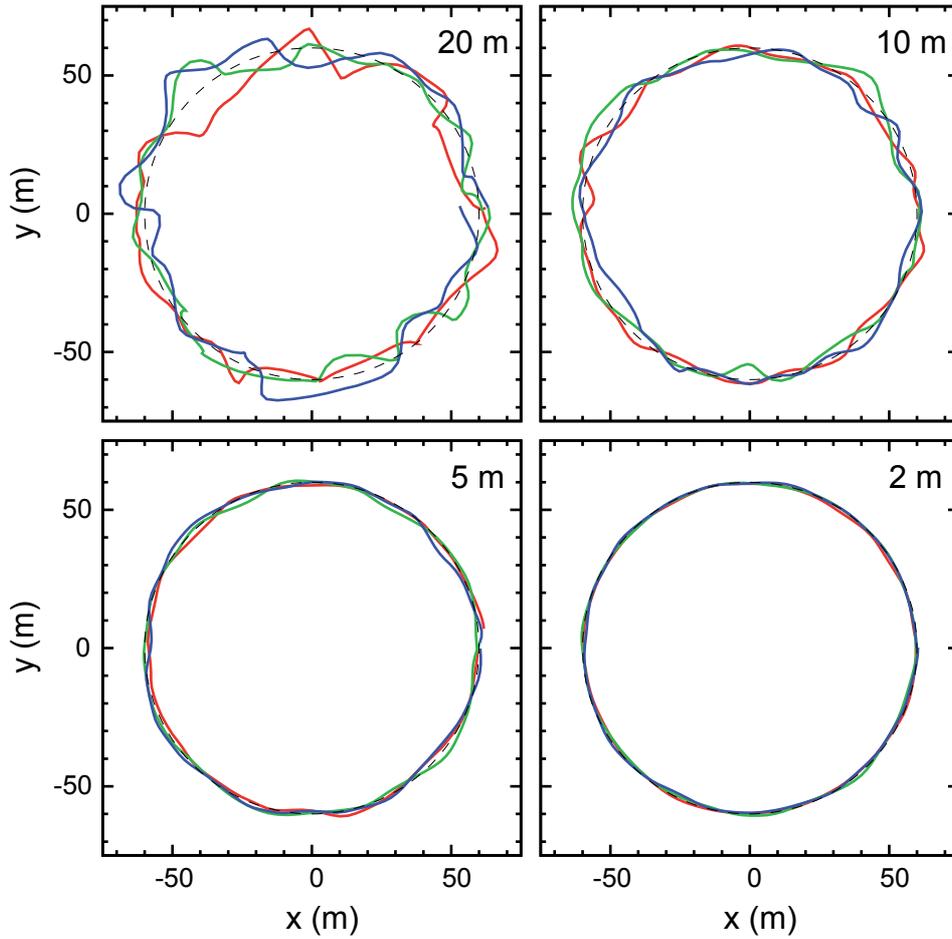
### Supplementary Figure 3

Adjacency matrix for the network shown in Figure 3a. Birds are ordered along both axes according to their rank in the overall hierarchy. The colours show the  $\bar{\tau}_{ij}$  value for each cell of the matrix, with reds corresponding to positive and blues to negative values. For those  $ij$  pairs that flew together only once (as well as for cases along the diagonal where  $i=j$ ), the corresponding cells of the matrix are marked with a cross. fi



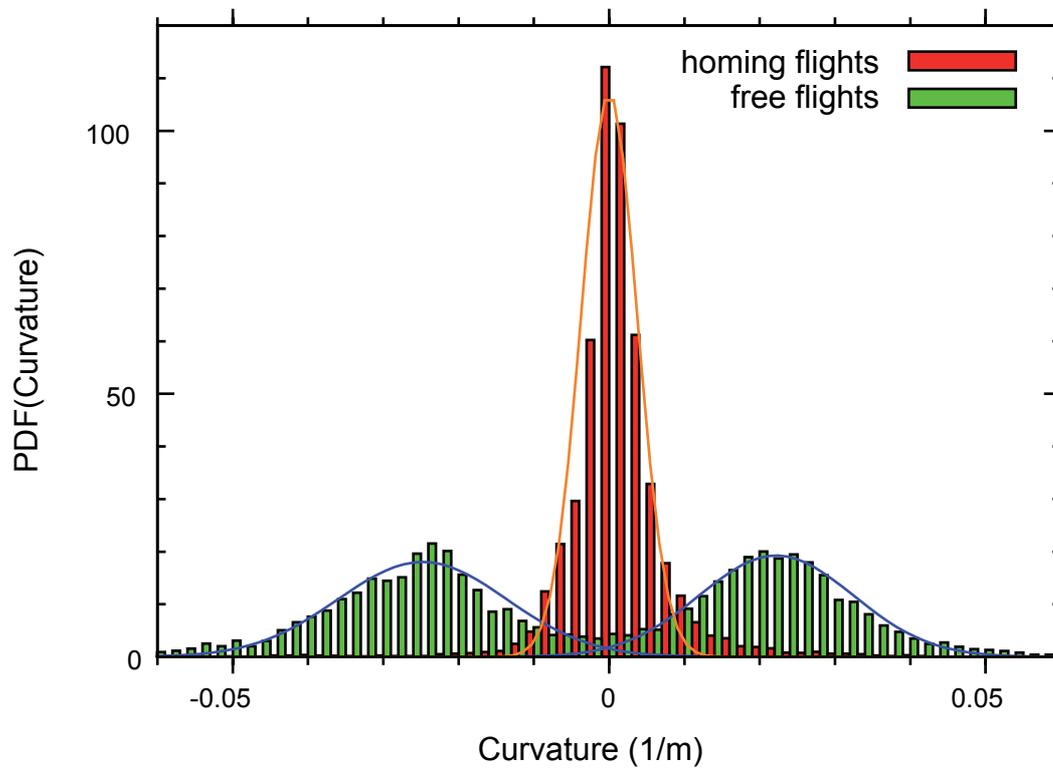
#### Supplementary Figure 4

Histograms of the directional correlation time values,  $\tau_i$ , across every flight for each subject. Positive values indicate leading (i.e., being copied in one's directional shifts by other flock members), while negative values indicate following (i.e., responding to directional shifts in other flock members' movement). The first nine birds (in red) correspond to those included in the overall network in Figure 3a, and are shown here in the order they appear in the hierarchy. The remaining four birds (in green) did not participate in sufficient numbers of flights to be included in the overall network. The histograms show that the average value  $\bar{\tau}_i$  provides a good summary measure of the distribution of  $\tau_i$  values.



### Supplementary Figure 5

Trajectories generated to simulate the effect of GPS positional error. Circular “flights” lasting 10 minutes were generated, with a radius of 60 m and a periodicity of 25 s, for a flock of individuals with a pre-defined hierarchical structure (see Supplementary Methods for further detail). The maximum radius of the noise superimposed onto each datapoint was 20, 10, 5 and 2 m (shown here in separate panels). For clarity, only one circle for each of three birds is shown at each magnitude of error tested, with the dashed line showing the original (no-noise-added) circle. Analysis of the effects of such errors on the robustness of our method for determining leader-follower relationships is summarised in Supplementary Table 2 and in the Supplementary Methods.



### Supplementary Figure 6

Histogram illustrating the distribution of turning directions in flocks' trajectories during free flights (green bars) and homing flights (red bars). The centre (curvature value of 0) corresponds to straight flight, while positive values indicate counter-clockwise (i.e. left) turns and negative values clockwise (right) turns. Thin blue and orange lines correspond to Gaussians fitted to the free and homing flight curvature distributions, respectively. See Supplementary Methods for further detail.

## 2. Supplementary Methods

### Calculation of directional correlation delay

To determine leader-follower relationships inside the flock we calculated the directional correlation delay for each pair of birds  $i$  and  $j$  ( $i, j = A..M, i \neq j$ ), and also for each bird  $i$  with regard to the average direction of motion of the rest of the flock. The directional correlation delay for a pair is  $C_{ij}(\tau) = \langle \vec{v}_i(t) \cdot \vec{v}_j(t + \tau) \rangle$ , where  $\vec{v}_i(t)$  is the normalised velocity of bird  $i$ . Note that  $C_{ij}(\tau) = C_{ji}(-\tau)$ . We then determined the maximum value of the  $C_{ij}(\tau)$  correlation function,  $\tau_{ij}^*$ , which we identify as the directional correlation delay time. Negative  $\tau_{ij}^*$  values mean that the flight direction of the  $i$ th bird falls behind that of the  $j$ th bird, and can thus be interpreted as a case of  $j$  leading. For every pair, we extracted from  $\tau_{ij}^* = -\tau_{ji}^*$  the positive value as a directed edge pointing from the leader to the follower.

The directional correlation function for a bird and a rest of the flock is  $C_i(\tau) = \langle \vec{v}_i(t) \cdot \vec{v}_{t,j}(t + \tau) \rangle$ , and we denote the maximum value of this function as  $\tau_i^*$ . To test for a link between leadership as defined by the correlations and the relative spatial position of birds inside the flock, we calculated for every pair the average projected distance onto the direction of motion of the whole flock as  $d_{ij} = \langle d_{ij}(t) \rangle_t$  and  $d_{ij}(t) = (\vec{x}_i(t) - \vec{x}_j(t)) \cdot \vec{v}_{flock}(t)$ , where  $\vec{x}_i(t)$  is the position of bird  $i$ , and  $\vec{v}_{flock}(t)$  is the normalised velocity of the whole flock  $\vec{v}_{flock}(t) = \langle \vec{x}_k(t) \rangle_k / \left| \langle \vec{x}_k(t) \rangle_k \right|$ .

Prior to analysis, all trajectory data collected was filtered according to two criteria. First, in order to minimise the effect of missing datapoints, we included only those points in the analysis for which the interval  $t-0.4s$  to  $t+0.4s$  contained at least two of the five possible positional fixes as logged by the GPS device. Second, for the calculation of the  $C_{ij}(\tau)$  correlation function, we included only those pairs of datapoints from birds  $i$  and  $j$  where the two birds were a maximum of 100 m apart (i.e.,  $d_{ij} < 100$  m).

### Validation of directional correlation delay method despite GPS error

As GPS data suffers from positional error (our independent tests showed that our device was accurate to approx. 1-2 m), as well as occasional missing datapoints, we generated sample data in order to test the robustness of our method based on the use of correlation functions. We generated trajectories (circular “flights” lasting 10 minutes, with a radius of 60 m and a periodicity of 25 s) that mimicked the track logs of a flock with a given hierarchical structure (i.e., 10 individual birds were programmed to respond to the directional changes in given flock members’ movement with a characteristic delay time: 0.1 s, 0.2 s, ... 0.9 s). We then superimposed positional perturbations of varying magnitude (white noise of maximally 2, 5,

10, 20, and 50 m) onto each datapoint, and simulated the correlation between successive datapoints by the application of a Gaussian filter ( $\sigma = 0.4$  s). The resulting paths (see Supplementary Figure 5) were then analysed using the directional correlation delay method outlined above. The results are summarised in Supplementary Table 2.

We found that even after introducing noise of up to 10 m radius around each datapoint, our method was able to detect the original hierarchy accurately, with errors in the  $\tau$  values averaging only 0.06-0.08 s. These errors are much smaller in magnitude than the average  $\tau$  values we obtained. Furthermore, the correlation between the hierarchical ranks of individual birds as entered into the simulation and that detected by our method has a Spearman correlation coefficient of 0.99 ( $p < 0.001$ ) for noise up to 10 m, and 0.83 ( $p = 0.003$ ) for noise up to 20 m (the latter is also the threshold above which directed loops first appear in the network), demonstrating that the method is highly robust in detecting the structure of the hierarchy, even in the presence of positional uncertainty. Only at 50 m maximal error do we fail to deduce a hierarchy that correlates with the originally defined ranks (and the networks obtained contain a large number of directed loops). Interestingly, once error rises to 20 m maximum, we no longer detect any edges at  $C_{\min} = 0.9$ , suggesting that edges that may be introduced erroneously due to positional uncertainty can be effectively filtered out at high  $C_{\min}$  threshold values. Hence, edges that we do detect in our data at  $C_{\min} = 0.9$  and  $C_{\min} = 0.99$  are extremely unlikely to represent false positives.

We next assessed the impact that missing datapoints may have had on our analysis, through generating trajectories at 5 Hz, and then removing either every second point deterministically, or the same number of datapoints but at random intervals along the track. At maximum positional errors of 10 m for each datapoint, both methods yielded an increase in our error in estimating  $\tau$  values (0.12 s with deterministic point-removal, and 0.16 s with random removal), nevertheless the correlation coefficients between the pre-set hierarchical ranks and those detected by the analysis remained very high at 0.96 and 0.94, respectively (both  $p < 0.001$ ). At smaller positional error (maximally 5 m), the removal of datapoints did not increase our error in estimating  $\tau$  values; correspondingly we are able to determine hierarchical ranks highly accurately (with correlation coefficients of 0.99 between pre-set and detected ranks,  $p < 0.001$  for both deterministic and random point-removal).

Finally, we also generated trajectories without any hierarchical structure, to assess the likelihood that our method would erroneously detect the existence of hierarchical networks in cases where there were none. For small positional errors (maximally 5 m),  $\tau$  values are reported as 0, and no edges are detected. For maximal error of 10 m or higher, we begin to detect edges (i.e.  $\tau$  values above 0), but these are accompanied by the appearance of loops, such that the resulting networks are not hierarchical. At  $C_{\min} = 0.9$ , no edges are detected, even at maximal error of 50 m, which suggests that the false  $\tau$  values obtained never have correlation coefficients above 0.9 (whereas we did obtain such – and stronger – correlations for edges in our data). These simulations with non-hierarchically organised flocks thus confirm that given an egalitarian system, our analysis is extremely unlikely to detect anything resembling the hierarchical networks we obtained from our data – confirming that we are unlikely to have made a type I error by favouring the existence of hierarchies over a null

hypothesis of egalitarian organisation. In other words, we can be confident that hierarchical organisation rather than egalitarianism better describes our data.

In sum, therefore, our method based on evaluating correlations between shifts in trajectories is able to reproduce, reliably, the underlying hierarchical structure of flock movement, even in the presence of error in recording the subjects' true position. Up to 10 m maximal noise, we are able to detect the original ranks highly accurately, with only small errors in  $\tau$  values, and are unlikely to introduce false edges (particularly in the case of edges with high  $C_{ij}(\tau)$  values). Given the estimated accuracy of our devices, the actual errors in our track logs are likely to be of a smaller magnitude, hence we can be reasonably certain that the hierarchical structures we observe are accurate representations of the existing interrelations within the flock.

### **Assessment of the distribution of left and right turning events during flock flights**

Our analysis of laterality effects (Table 1 of the main text) rests on the assumption that there were no biases in the propensity of flocks to make turns either to the left or to the right. Such biases could potentially lead to an artefactual association between  $Q_{left}$  and  $Q_{forward}$  – given that follower birds tend to be located behind leaders (Fig. 3b), if the flock turns more often in one direction than the other, then these follower birds will correspondingly tend to be on one side of the leaders than the other simply as a result of changes in the arrangement of individuals as the flock turns. Therefore, to examine whether any biases were evident in flocks' choices to make left or right turns, we calculated the density distribution of trajectory curvature values throughout each free and homing flight performed by our flocks.

For the purposes of the present analysis, the “flock” was defined as a group of at least five birds whose inter-individual distances did not exceed 30 m. The trajectory of the centre of mass of the flock was smoothed with a Gaussian filter ( $\sigma = 0.4$  s). Curvature was then calculated as  $\kappa(t) = |\dot{r}(t) \times \ddot{r}(t)| / |\dot{r}(t)|^3$ , where  $\dot{r}(t)$  and  $\ddot{r}(t)$  correspond to the first and second derivatives of  $r(t)$ , respectively,  $\times$  is the cross product, and  $r(t)$  denotes the trajectory of the centre of mass of the flock in the horizontal plane. Events such as a bird leaving or joining the group or one or more of the GPS loggers losing their satellite signal caused a jump in the trajectory of the centre of mass of the flock, which led to a discontinuity in the  $\kappa(t)$  function. To omit these points we used for the curvature statistics only those  $\kappa(t)$  values for which the flock contained the same individuals for the interval  $[t - 3 \text{ s}, t + 3 \text{ s}]$ .

Supplementary Figure 6 shows the probability density functions we thus obtained for the distribution of curvature values throughout all homing and free flights. For homing flights, the distribution is bimodal and symmetrical on either side of zero (where zero represents straight, non-curving flight), meaning that flocks flew curved trajectories most of the time, but these were no more likely to tend towards the left than towards the right. The mean ( $\pm$  SD) of the curvature for counter-clockwise (i.e. left) turns was  $0.023 \pm 0.011$  1/m, while for clockwise (right) turns it was  $-0.027 \pm 0.013$  1/m (meaning that the typical radius of circles flown during free flights was  $\sim 40$  m). Gaussians fitted to these data had parameter values  $\mu = 0.022$  and  $\sigma = 0.010$ , and  $\mu = -0.025$  and  $\sigma = 0.011$ , respectively.

During homing flights the distribution of curvature measurements was unimodal, symmetrical, and centred tightly around zero (mean  $\pm$  SD:  $0.0002 \pm 0.007$  1/m; fitted Gaussian  $\mu = 0.000$  and  $\sigma = 0.004$ ) meaning that the majority of the time flocks flew straight or close-to-straight trajectories and when they did not, they were no more likely to perform left- than right-curved turns. This held true for data pooled across all flights, as shown in Supplementary Figure 6, as well as for each of the four homing flights when they were analysed separately.

Taken together these results suggest that during both free and homing flights, flocks showed no tendency to perform more left- than right-handed turns. This in turn implies that we are unlikely to be dealing with any introduced bias in our analysis of laterality effects due to the spatial dynamics of flocks during turns – instead the results (especially when considered in conjunction with the  $\tau_{left} - \tau_{right}$  analysis also in Table 1) likely reflect a left-eye/right-hemispheric advantage in the processing of social information.

### 3. Supplementary Tables

#### Supplementary Table 1

Summary of hierarchical structure observed in each free and homing flight performed by subjects in flocks.

Flight	$C_{\min} = 0.5$ *					$C_{\min} = 0.9$				
	# nodes	# links	# loops	$Q_{\text{hier}}$ **	$p_{\text{rand}}$ ***	# nodes	# links	# loops	$Q_{\text{hier}}$	$p_{\text{rand}}$
<b>Free flights</b>										
<b>FF1</b>	7	17	0	1	0.027	3	2	0	1	-
<b>FF2</b>	9	27	0	1	0.001	9	21	0	1	0.017
<b>FF3</b>	6	10	0	1	-	3	3	0	1	-
<b>FF4</b>	6	10	1	0.9	-	2	1	0	1	-
<b>FF5</b>	9	37	0	1	< 0.001	9	21	0	1	0.020
<b>FF6</b>	7	9	0	1	-	4	2	0	1	-
<b>FF7</b>	4	5	0	1	-	3	2	0	1	-
<b>FF8</b>	6	6	0	1	-	6	6	0	1	-
<b>FF9</b>	7	20	0	1	0.007	6	7	0	1	-
<b>FF10</b>	8	17	0	1	0.051	3	1	0	1	-
<b>FF11</b>	9	31	5	0.9	0.002	8	15	2	0.87	-
<b>Homing flights</b>										
<b>HF1</b>	9	34	7	0.97	0.001	9	34	7	0.97	0.001
<b>HF2</b>	8	29	0	1	< 0.001	8	29	0	1	< 0.001
<b>HF3</b>	7	16	0	1	0.041	7	16	0	1	0.036
<b>HF4</b>	8	28	0	1	< 0.001	8	28	0	1	< 0.001

FF = free flight; HF = homing flight.

\*  $C_{\min}$  = the threshold value for the  $C_{ij}(\tau)$  correlation. For any given  $C_{\min}$ , only those edges (links) that exceed the given minimum value are used in generating the network.

\*\*  $Q_{\text{hier}}$  = the proportion of the total number of edges that point in the same direction within the best possible hierarchical layout of the network. In a randomised network this value is around 0.5.

\*\*\* Statistical significance when the number of directed loops in the network is compared to randomised data. Randomised data were based on the Erdős-Rényi model for directed random networks. The p-value (denoted as  $p_{\text{rand}}$ ) refers to the probability of obtaining networks with as many or fewer loops as the total we observed, given the number of nodes and edges (but not their direction) in the network.  $p_{\text{rand}}$  is calculated only for those networks that contain at least twice as many links as they contain nodes, since the probability of a lack of directed loops in sparse graphs is particularly high.

## Supplementary Table 2

Results of directional correlation delay analysis on trajectories generated to mimic the track logs of a flock with a pre-defined hierarchical structure (see Supplementary Methods). The effect of GPS positional error and of omitted datapoints is examined, along with the likelihood that the method falsely detects hierarchical structures in their absence.

$R_{noise}^{max}$ (m)	$R_{noise}^{eff}$ (m)	missing time points	For $ij$ pairs		For each $i$		$\tau$ -Network			Spearman corr.		
			$\langle C_{ij}^* \rangle$	Dev. (s)	$\langle C_i^* \rangle$	Dev. (s)	# Loop	# Link	# Links	r	p	
			$C_{min} = 0$						$C_{min} = 0.9$	$C_{min} = 0$		
			<b>Hierarchically generated trajectories:</b>						$\tau_{ij} = (i - j) \cdot 0.1$			
<b>2</b>	<b>0.54</b>	<b>No</b>	0.998	0.06	0.995	0.06	0	41.5	41.5	0.99	< 0.001	
<b>5</b>	<b>1.4</b>	<b>No</b>	0.99	0.06	0.98	0.06	0	40	40	0.99	< 0.001	
<b>5</b>	<b>1.4</b>	<b>Determ.</b>	0.98	0.06	0.97	0.06	0	40.3	40.3	0.99	< 0.001	
<b>5</b>	<b>1.4</b>	<b>Random</b>	0.97	0.06	0.97	0.06	0	40.6	40.6	0.99	< 0.001	
<b>10</b>	<b>2.7</b>	<b>No</b>	0.94	0.08	0.94	0.06	0	39.1	39.1	0.99	< 0.001	
<b>10</b>	<b>2.7</b>	<b>Determ.</b>	0.90	0.12	0.89	0.07	0.1	38.2	9.1	0.96	< 0.001	
<b>10</b>	<b>2.7</b>	<b>Random</b>	0.90	0.16	0.89	0.08	2.6	38.5	6.6	0.94	< 0.001	
<b>20</b>	<b>5.4</b>	<b>No</b>	0.74	0.3	0.73	0.11	12.4	36.6	0	0.83	0.003	
<b>50</b>	<b>14</b>	<b>No</b>	0.24	0.6	0.23	0.4	21	39.4	0	0.57	0.094	
			<b>Non-hierarchically generated trajectories:</b>						$\tau_{ij} = 0$			
<b>2</b>	<b>0.54</b>	<b>No</b>	0.998	0	0.998	0	0	0	0	-	-	
<b>5</b>	<b>1.4</b>	<b>No</b>	0.99	0	0.99	0	0	0	0	-	-	
<b>10</b>	<b>2.7</b>	<b>No</b>	0.94	0.06	0.94	0	0.8	13.1	0	-	-	
<b>20</b>	<b>5.4</b>	<b>No</b>	0.73	0.3	0.73	0.12	17.7	35.5	0	-	-	
<b>50</b>	<b>14</b>	<b>No</b>	0.24	0.4	0.23	0.38	16.6	39.2	0	-	-	

$R_{noise}^{max}$  = radius representing the maximum possible size of the noise (error) assigned to each datapoint of the generated trajectory.

$R_{noise}^{eff}$  = average deviation of the trajectory from the originally generated circular path after the addition of  $R_{noise}^{max}$  noise and Gaussian averaging.

$C_{min}$  = the threshold value for the  $C_{ij}(\tau)$  correlation. For any given  $C_{min}$ , only those edges (links) that exceed the given minimum value are used in generating the network.

missing time points = indicates whether any datapoints were excluded from the generated trajectories before analysis, in order to simulate the effect of missing GPS data. Removal was performed either in a deterministic manner (“Determ.”; every second datapoint removed) or randomly (“Random”; randomly selected half of datapoints removed).

$\langle C_{ij}^* \rangle$  = mean value of the maximum of the directional correlation functions for the pair

$\langle C_i^* \rangle$  = mean value of the maximum of the directional correlation functions for each bird

Dev. = absolute value of the difference between the  $\tau$  values entered into the simulation and those observed following the relevant manipulation (addition of positional error or the removal of datapoints).

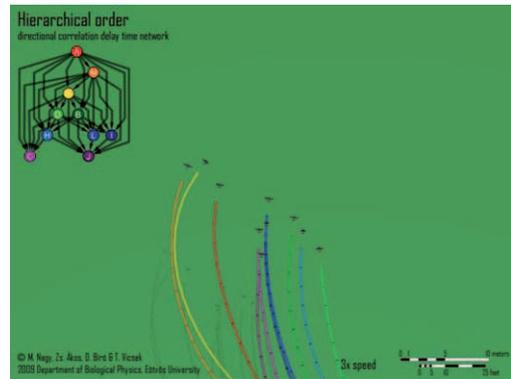
$\tau$ -Network = mean number of directed loops and links in the network composed from the  $\tau_{ij}^*$  values as directed edges (links).

Spearman Corr. = Spearman correlation coefficient (r) and significance level (p) between the originally defined order and that deduced from the  $\tau$ -network obtained (n = 10).

## 7. Supplementary Movie Legends

### Supplementary Movie 1

Animation showing a 100-second segment of a free flight performed by a flock of 10 pigeons, reconstructed from GPS data. The segment is taken from the same flight as that depicted in Figure 2, and plays at 3x actual speed. Individuals are coloured according to their rank in the hierarchy (shown in the top left), where ranks are determined on the basis of pairwise directional correlation delay times (see Figure 2 and explanation in main text). Birds shown in colours near the red end of the spectrum are higher in the hierarchy, while those lower down are shown in blues and purples. A longer version of the animation is available at <http://hal.elte.hu/pigeonflocks/>.



### Supplementary Movie 2

Animation showing a 100-second segment of a homing flight performed by a flock of nine pigeons, reconstructed from GPS data (playback at 3x actual speed). Individuals are coloured according to their rank in the hierarchy (shown in the top left), where ranks are determined on the basis of pairwise directional correlation delay times (see Figure 2 and explanation in main text). Birds shown in colours near the red end of the spectrum are higher in the hierarchy, while those lower down are shown in blues and purples. Inset in bottom left shows an overview of the animated segment of the homing paths (see Supplementary Figure 2b for the full journey), with a moving cursor indicating the flock's progress. The x and y axes show distance from home, where home is at coordinates 0,0. Bird G (shown in purple) later split from the group and was the last to return home. A longer version of the animation is available at <http://hal.elte.hu/pigeonflocks/>.

