Peering into the world of wild passerines with 3D-SOCS: synchronized video capture and posture estimation

Michael Chimento†[1,2,3], Alex Hoi Hang Chan†[1,4], Lucy M. Aplin[2,3,5], and Fumihiro Kano[1,4]

[1] *Centre for the Advanced Study of Collective Behaviour, University of Konstanz; Konstanz, Germany*
[2] *Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland*
[3] *Cognitive and Cultural Ecology Research Group, Max Planck Institute of Animal Behavior; Radolfzell, Germany*
[4] *Department of Collective Behaviour, Max Planck Institute of Animal Behavior; Radolfzell, Germany*
[5] *Division of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, Australia*

June 30, 2024

†These authors contributed equally to the manuscript and share first, corresponding authorship. Inquiries can be addressed to mchimento[at]ab.mpg.de or hoi-hang.chan[at]uni-konstanz.de.

Abstract

Collection of large behavioral data-sets on wild animals in natural habitats is vital in ecology and evolution studies. Recent progress in machine learning and computer vision, combined with inexpensive microcomputers, have unlocked a new frontier of fine-scale markerless measurements. Here, we leverage these advancements to develop a 3D Synchronized Outdoor Camera System (3D-SOCS): an inexpensive, mobile and automated method for collecting behavioral data on wild animals using synchronized video frames from Raspberry Pi controlled cameras. Accuracy tests demonstrate 3D-SOCS' markerless tracking can estimate postures with a 3mm tolerance. To illustrate its research potential, we place 3D-SOCS in the field and conduct a stimulus presentation experiment. We estimate 3D postures and trajectories for multiple individuals of different bird species, and use this data to characterize the visual field configuration of wild great tits (*Parus major*), a model species in behavioral ecology. We find their optic axes at approximately $\pm 60^\circ$ azimuth and -5° elevation. Furthermore, birds exhibit individual differences in lateralization. We also show that birds' convex hulls predicts body weight, highlighting 3D-SOCS' potential for non-invasive population monitoring. In summary, 3D-SOCS is a first-of-its-kind camera system for wild research, presenting exciting potential to measure fine-scaled behaviour and morphology in wild birds.

1 Introduction

Automated, large-scale data collection methods to measure animals' behaviour in their natural habitat are of growing importance for understanding social systems [1, 2], migration and dispersal [3–5], disease and information transmission [6–8] responses to climate and weather events [9, 10], and conservation impacts [11–13]. Increasingly, behaviour in the wild is measured using animal-borne sensor technologies, including GPS and Passive Integrative Transponders (PIT). GPS or accelerometers are employed to measure large scale movement patterns and fine-scale behaviors [14], but are often limited by the size, battery life and data storage of the device [15]. To follow smaller animals or to increase the scope of studies, PIT tags that utilize Radio Frequency Identification (RFID) technology are commonly used. Thanks to its small size and durability, RFID technology is widely employed to monitor wild bird and rodent populations [16–19]. Analyzing co-occurrence data from RFID tags has provided insights into social interactions, such as parental provisioning [20] and social networks [1, 21, 22]. Additionally, the widespread adoption of low-cost single-board computers, such as Raspberry Pi [23], has further enabled the use of RFID technology in experimental devices designed to study cognition [7, 24–26], or to manipulate social associations in wild populations [27, 28].

While RFID technology provides information about the presence and absence of individuals near RFID antennas, it does not provide detailed information about their behaviors. Furthermore, RFID readings may be affected by a variety of environmental conditions such as humidity. An emerging method that has the potential to address this limitation involves the use of computer vision and machine learning tools [29–31], a development facilitated by a concomitant increase in the volume and resolution of image data [32, 33]. Computer vision enables the automatic quantification of fine-scaled behaviors from images, not only increasing the volume of behavioral data but also enhancing the objectivity of behavior coding. Although initial applications were primarily in controlled laboratory settings [34–46], computer vision is also being applied in the field. It has been used to monitor animals from drones [47–49], quantify courtship displays [50], analyze parental care behaviors [51], study underwater interactions [52], identify individuals [53, 54], and more.

A particular branch of computer vision that is revolutionizing the way we measure animal behavior is 2D and 3D posture estimation [11, 31, 55]. Posture estimation refers to the task of estimating the position of morphological keypoints (e.g eyes, limbs, joints) from an image. With the recent development of frameworks like DeepLabCut, DeepPoseKit and SLEAP, multi-animal posture estimation in 2D and subsequent automated behavioural analysis have become possible [40, 46, 48, 56]. In 3D, frameworks like Anipose, DANNCE and 3D-MuPPET [37, 41, 45] have also been developed to estimate postures in 3D using multiple camera views (but see [44] for single view 3D estimation). However, these frameworks are often limited to tracking a single animal, and only a few recent studies have demonstrated the ability to estimate the 3D postures of multiple individuals simultaneously (see [37, 57, 58]). Furthermore, to date, applications in the wild have been limited, with notable exceptions including posture estimation of a single cheetah [59] and of three pigeons in semi-captive outdoor systems [37]. The lack of adoption of 3D animal posture tracking can be attributed to a few main challenges. Firstly, algorithms for multiple individuals are less mature due to correspondence matching (identifying and matching the same points across different frames or views) and object trajectory tracking [60]. Secondly, 3D tracking often requires multiple synchronized camera views, which is either expensive or difficult to set up outside of the lab [50].

Yet, the potential benefits of applying 2D and 3D posture tracking in the wild are immense. Automatic posture estimation is not only less labor-intensive but also achieves greater objectivity than manual behavior coding based on ethograms [61]. Classification of posture-based measurements can provide direct information about individual and group behaviors such as social and dominance interactions [60], decision making and information transfer [47, 62] and courtship behaviour [50, 63]. Marker-based systems have been recently employed to measure the 3D head orientation of birds, presenting an effective method for estimating visual field use, lateralization of eye use, and attentional foci of birds [35, 62, 64–66]. However, marker-based systems are limited to the laboratory, highlighting the need for the development of markerless systems that could extend this approach to field setups.

Here, we introduce 3D synchronized outdoor camera system (3D-SOCS), a mobile, flexible and low-cost system that combines computer vision and sensor technology to measure both the proximity and postures of wild animals (Figure 1). With this system, we are able to track the 3D posture of individuals from a population of great tits (*Parus major*) and blue tits (*Cyanistes caeruleus*) within a 3mm accuracy. To showcase the capabilities of 3D-SOCS, we conducted a stimulus presentation experiment to examine the use of visual fields and lateralization of eye use in great tits, based on the tracking of their head orientation in 3D [64–66]. While our system typically uses six cameras, we demonstrate that 3D-SOCS can operate effectively with just two cameras as a minimal setup. We also provide detailed documentation and code to enable other researchers to deploy the system (see section 5).

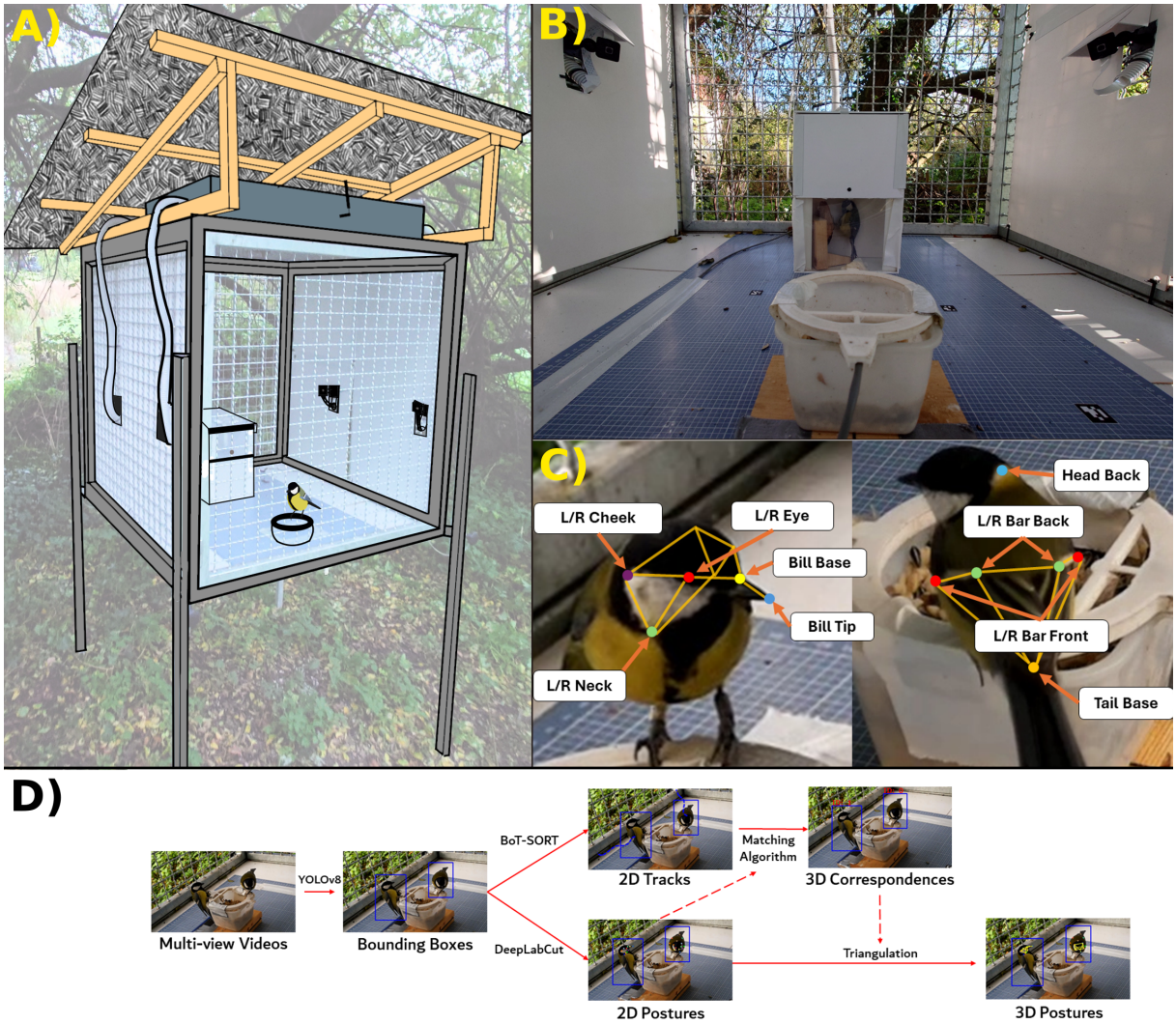

Figure 1: **Overview of 3D-SOCS**. A) Our mobile data collection system was housed in an elevated 1x1m cage with a roof for protection from the elements. The cage was designed to exclude large non-target species but to allow easy access for great tits. We installed white PVC plate walls to remove possible distractions in their lateral visual field. B) Whenever a tagged bird landed on the feeding perch antenna (foreground), the PLDC glass of the stimulus presentation device (background) turned transparent after a 1s delay, revealing the stimulus. Six frame-synchronized cameras placed around the cage filmed birds from various angles as they regarded the stimulus. C) We extracted keypoints on the head and body of birds from each frame, and used this information to create a 3D representation of the bird. D) Summary of the 3D tracking pipeline applied to videos.

2 Results

2.1 3D-SOCS

3D Synchronized Outdoor Camera System (3D-SOCS) is a versatile markerless 3D posture tracking system for small passerine birds that runs on a battery and can be readily deployed in the field (Figure 1A). We designed the system to be: 1) scalable, in terms of the number of cameras and units that can be deployed; 2) flexible, in terms of the type of experimental apparatus that it be deployed with, and 3) affordable (< 2500€), as the system is based on relative cheap micro-computers. See supplementary methods section S1.1 for detailed description of the hardware and software required to deploy and collect data with 3D-SOCS. While 3D-SOCS can use any rigid frame to mount the cameras, we used a 1x1m metal cage that housed our cameras and experimental devices (Figure 1B). We mounted six frame-synchronized cameras at various angles focused on a single feeder, each controlled by a Raspberry Pi Compute Module 4 (CM4), which achieved sub-millisecond frame synchronization using precision time protocol (PTP) at 30 frames per seconds.

Utilizing these synchronized cameras, we could reconstruct the 3D postures of birds visiting the feeder, and we were able to track multiple birds at the same time. Our target species was great tits, although a smaller number of tagged blue tits were also present, so we extended the model to detect their keypoints. We adopted a top-down approach similar to [37] by: 1) localizing individual birds in each frame using YOLOv8 [67], 2) estimating 14 unique 2D keypoints (Figure 1C) with DeepLabCut [40, 68], and 3) triangulating 2D postures into 3D (Figure 1D). For each unique individual recorded in the study ($N = 57$ great tits, $N = 20$ blue tits), we collated all 3D information and calculated the median head and body skeleton, then ensured the detections fit the median head skeletons by computing the estimated rotation translation of the skeleton in 3D space. This approach leverages the assumption that certain keypoints on the head of a bird form a rigid body (see [38]), allowing for missing keypoints to be filled and relative distances between keypoints to be fixed, leading to more consistent detections. We refer to section 4.6 for the detailed 3D tracking pipeline and supplementary video S1 for qualitative results.

2.2 System accuracy

We first evaluated the synchronization performance of 3D-SOCS, by comparing the time offset for the frame capture time for each camera compared to the mean time across all 6 cameras. The frames captured using the system were highly synchronized, with a mean offset of $1.15 \times 10^{-5} ms$ and 99% intervals of -$22ms - 28ms$, within the frame-by-frame difference of $33ms$ at 30fps. We then evaluated 3D-SOCS to ensure all measured keypoint estimates of wild birds were accurate throughout the experimental period. Overall, we achieved low median 2D detection error and 3D reprojection error of 0.24px and 4.61px respectively, and an average of 18% frame loss rate across all keypoints (Tables S3, S5). Body keypoints were lost more frequently than head keypoints (Head: 13%, Body: 28%). We also explored the minimal setup required for 3D posture tracking using 3D-SOCS by using 2 cameras for triangulation (see section 4.6). This approach yielded similarly accurate results, but with a higher frame loss rate, especially for body keypoints (Table S4).

We further validated the overall accuracy of the 3D-SOCS by comparing 3D-SOCS' estimates with estimates from a marker-based motion capture system with sub-millimeter precision (SMART-BARN [35]), which we treated as the ground truth. 3D-SOCS' measurements for head keypoint locations were normally distributed around 0 (visualized in Figure S4). The upper 95% highest density interval (HDI) of keypoint detections were no further than 3mm from their ground-truth position, with a median deviation of 1.05mm. Distributions of absolute rotation errors for yaw, pitch, and roll were also overall distributed around 0, with an upper 95% HDI between 7° and 11° depending on the rotation type (Table 1). We tested two different camera resolutions corresponding to narrow (1920x1080px) and wide (1640x1280px) fields of view. While 1080p resolution resulted in slightly better estimations of yaw and pitch, the two resolutions were generally comparable. In summary, our findings show no evidence of systematic deviations from the ground-truth values in the measurements.

2.3 Estimating visual field use

As a case study to test the capabilities of 3D-SOCS, we set out to determine the visual field use of great tits. For 5 days we deployed a stimulus presentation box and a feeder with an RFID antenna, which could read unique identities for birds wearing a passive-integrated transponder (PIT) leg band (Figure 1). When a PIT-tagged bird landed on the reader, an opaque PDLC film on the stimulus device was triggered after a 1s delay, which turned the film transparent and revealed the stimulus. We subsequently measured the head rotation angles relative to the stimulus within 3 seconds of the stimulus presentation to determine the areas

Resolution	Overall Error (mm)	Yaw Error (°)	Pitch Error (°)	Roll Error (°)
Overall	1.05 (2.81)	1.92 (7.25)	2.35 (10.52)	2.13(7.48)
1920x1080px	1.01 (2.95)	1.57 (5.16)	2.57 (9.30)	2.04 (7.69)
1640x1280px	1.15 (2.57)	2.34 (8.53)	2.17(11.53)	2.33(7.46)

Table 1: Summary table of system accuracy tests with six cameras. Median absolute errors and upper 95% highest density interval (HDI) in parenthesis. Values of head keypoint positions and roll, pitch, yaw angles measured by 3D-SOCS were compared to SMART-BARN as ground truth. A value of zero would indicate a measurement that matched ground truth. Keypoints estimated from six cameras, recorded at either a resolution of 1920x1080px (narrow FOV) or 1640x1280px (wide FOV).

of the visual fields most frequently used by birds. This area should encompass the optic axis they use, which is situated close to the birds' fovea or area *centralis*. The width of this area is likely influenced by the degree of eye movement as well as by measurement errors [64, 69–71].

We presented four different visual stimuli: 1) an empty box as a control for the clicking sound of the stimulus box, 2) food stimulus presenting a tray of mealworms (a highly desirable food item), 3) food video stimulus presenting a screen displaying a looped video of a single mealworm wiggling in place, and 4) social stimulus presenting a taxidermy female great tit. In total, we estimated head posture data of $34,587$ frames from $N = 57$ great tits (26 females, 29 juveniles). Each bird was recorded as visiting 3D-SOCS on average 18 times (range: 1-62). The orienting responses to the visual stimulus were defined as moments when the bird's head was stationary (head-fixation); frames in which the bird exhibited rapid saccade-like head movements were excluded from the analysis.

We took several complementary approaches to assessing birds' visual field use. Figure 2D summarizes head azimuth and elevation angles relative to all stimuli using a heatmap. Usage of both lateral and binocular fields are clearly shown by the high density regions. The hot-spots in the lateral field indicate the most likely positions of the optic axis [64, 66]. When binning the marginal distributions of azimuth and elevation into $5°$ bins, the most observations fell between $\pm 60°$ and $\pm 65°$ azimuth and between $-5°$ and $-10°$ elevation, indicating the approximate locations of foveal projections.

In order to statistically validate this estimate, we used a Bayesian mixture model that estimated a 2-dimensional normal distribution for each eye, composed of an azimuth and an elevation component, and a covariance matrix between azimuth and elevation parameters. A mixture parameter estimated the proportional contribution of each eye's distribution to the overall distribution, although we note that this could not account for variation in eye preferences between individuals. Model results are reported in table 2A, and fully summarized in table S6. Visual field use was estimated to be most variable under the control condition, where no specific stimulus was present to observe. Across stimuli, we found a range of azimuth estimates between $-63°$ to $-53°$ for the left eye and $56°$ to $70°$ for the right eye. While mean estimates varied between stimuli, the mean absolute value of azimuth estimates was $59.63°$, supporting our initial assessment of the optic axis at approximately $60°$. The slightly wider right eye azimuth estimate for the taxidermy bird stimulus was either caused by 1) birds attending to the wooden stand the bird was attached to, rather than the bird itself, or 2) eye movement as birds held their heads more toward the direction of the exit for the taxidermy bird. The elevation was estimated to be higher than our initial assessment, although this was because elevation was modeled as a normal distribution, yet its marginal distribution was positively skewed. In light of the rotation error measured in section 2.2, we conclude the optic axes' azimuth is approximately $60 \pm 10°$ and elevation $-5 \pm 10°$.

In order to analyze bird's preferences for using either eye, we fit a binomial model where the use of the lateral field of the right eye versus that of the left eye was predicted by varying effects for stimulus type and individual identity (Table 2B, fully summarized in Table S7). Age and sex showed no effect on lateralization, and these were excluded from the final model. We found more variation between birds than between stimuli (individual SD and 95% HPDI: $1.71(1.33, 2.2)$, stimuli: $0.81(0.31, 1.82)$. Interestingly, rather than a bimodal distribution of preferences, birds fell along a spectrum of lateralization (Figure S8). When controlling for differences between individuals, the taxidermy bird was estimated to elicit the strongest preference for the right eye ($\hat{P}(\text{use right lateral field}) = .75$). Weaker preference for the right eye was found in the control and food stimuli conditions, although the HPDI of these estimated probabilities crossed $.5$ indicating no strong preference.

Finally, to examine the temporal changes of birds' orienting responses, we plotted the density of head angles from the moment of stimulus presentation as heatmaps. To do this, we subset the data to observations with head elevation between -15 and 5 degrees (informed by the mixture model and system accuracy tests) to exclude observations where birds were likely not attending to the stimulus. This analysis provided a more

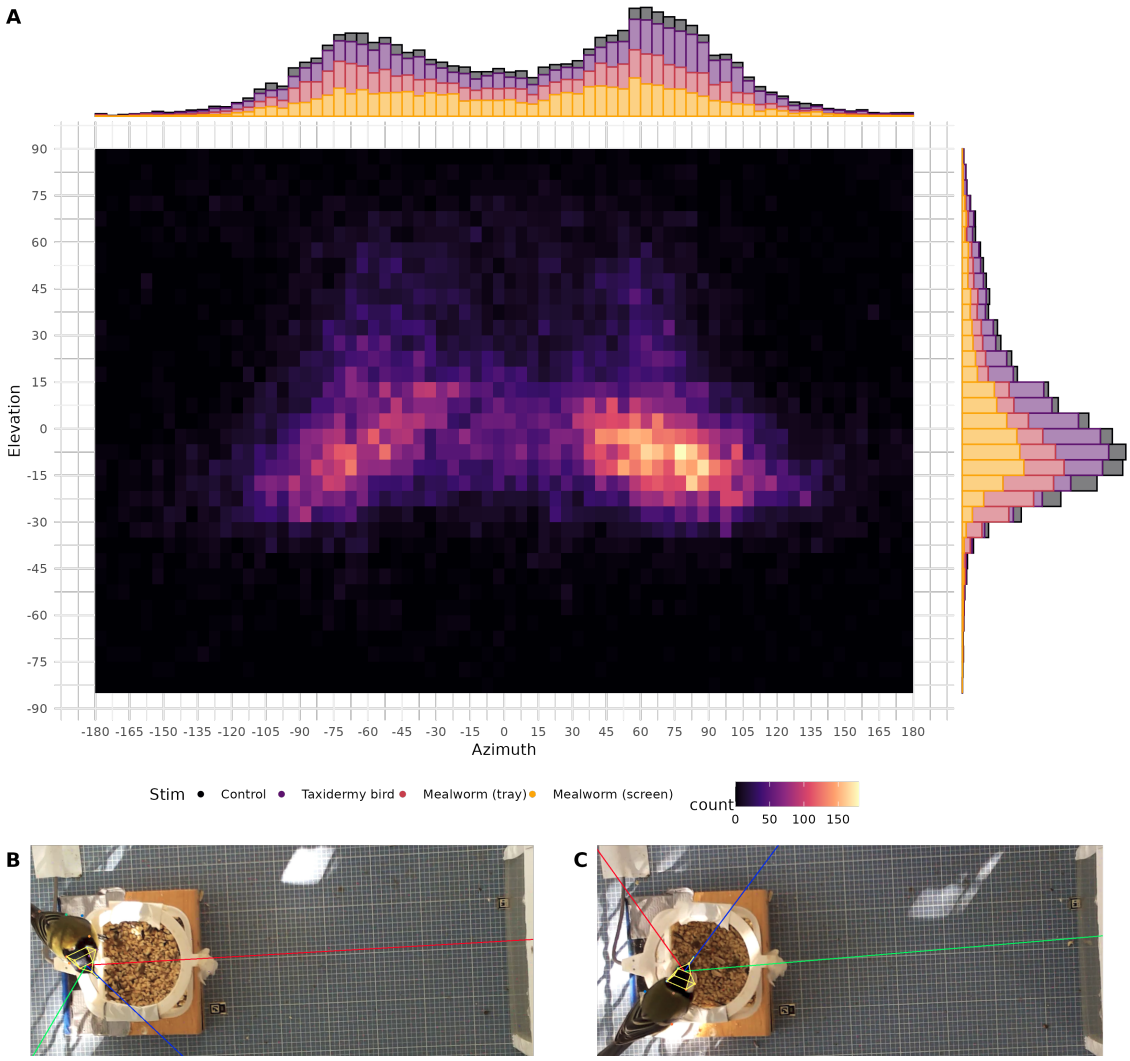

Figure 2: **Summary of great tits' visual field use.** A) A heat-map illustrating the count of observations of head azimuth (x-axis, degrees) and elevation (y-axis, degrees) for all four stimuli. Histograms are included as marginal plots. Birds particularly used areas around $\pm 60^\circ$ azimuth and -5° elevation to regard the stimuli. These areas likely encompasses their optic axes, as illustrated by frames taken when birds are regarding the stimulus with their **B)** left eye (red line rendered at -60° azimuth) and **C)** right eye (green line rendered at 60° azimuth).

A) Mixture model estimates *condition*	*left eye value*	*right eye value*	
azimuth mean (degrees)			
control	-66.01 (-69.35,-62.2)	52.63 (49.14,56.3)	
taxidermy bird	-59.24 (-60.88,-57.58)	72.2 (70.99,73.37)	
mealworm (tray)	-58.28 (-59.66,-56.89)	57.36 (55.94,58.76)	
mealworm (screen)	-54.87 (-56.36,-53.33)	56.42 (55.03,57.79)	
azimuth SD (degrees)			
control	37.46 (35.19,40.11)	48.21 (45.71,50.68)	
taxidermy bird	35.98 (34.9,37.09)	34.85 (33.8,35.94)	
mealworm (tray)	33.44 (32.42,34.52)	35.23 (34.18,36.31)	
mealworm (screen)	33.33 (32.3,34.38)	36.7 (35.65,37.8)	
elevation (degrees)			
elevation mean			1.97 (1.72,2.23)
elevation SD			24.16 (23.98,24.34)

B) Lateralization model estimates *condition*	*value*
P(prefer right eye)	
control	0.57 (0.45 0.70)
taxidermy bird	0.75 (0.65 0.84)
tray	0.54 (0.43 0.66)
screen	0.61 (0.50 0.72)
Varying effects (logit scale)	
SD (stimulus)	0.81 (0.31,1.82)
SD (ID)	1.71 (1.33,2.2)

Table 2: **Summary of model estimates**. Mean and 95% HPDI presented for all parameters. Models fit to data recorded within 3s after stimulus was presented. **A)** Bayesian mixture model described the estimated location of optic axes. **B)** Bayesian logistic regression with varying effects for stimulus type and bird identity was used to estimate differences in preferences for using left and right eyes when regarding stimuli.

nuanced picture of the birds' visual behavior over the time spent on the feeder (Figure 3). The clicking noise in the control condition appeared to elicit orienting responses, although this response was less structured compared to other stimuli and failed to maintain the birds' attention for as long. When shown real mealworm stimuli, birds were more likely to use their binocular fields. In contrast, they primarily used their right eye to regard both the mealworm video stimulus and the taxidermy bird stimulus. They also regarded the taxidermy bird at a slightly wider angle.

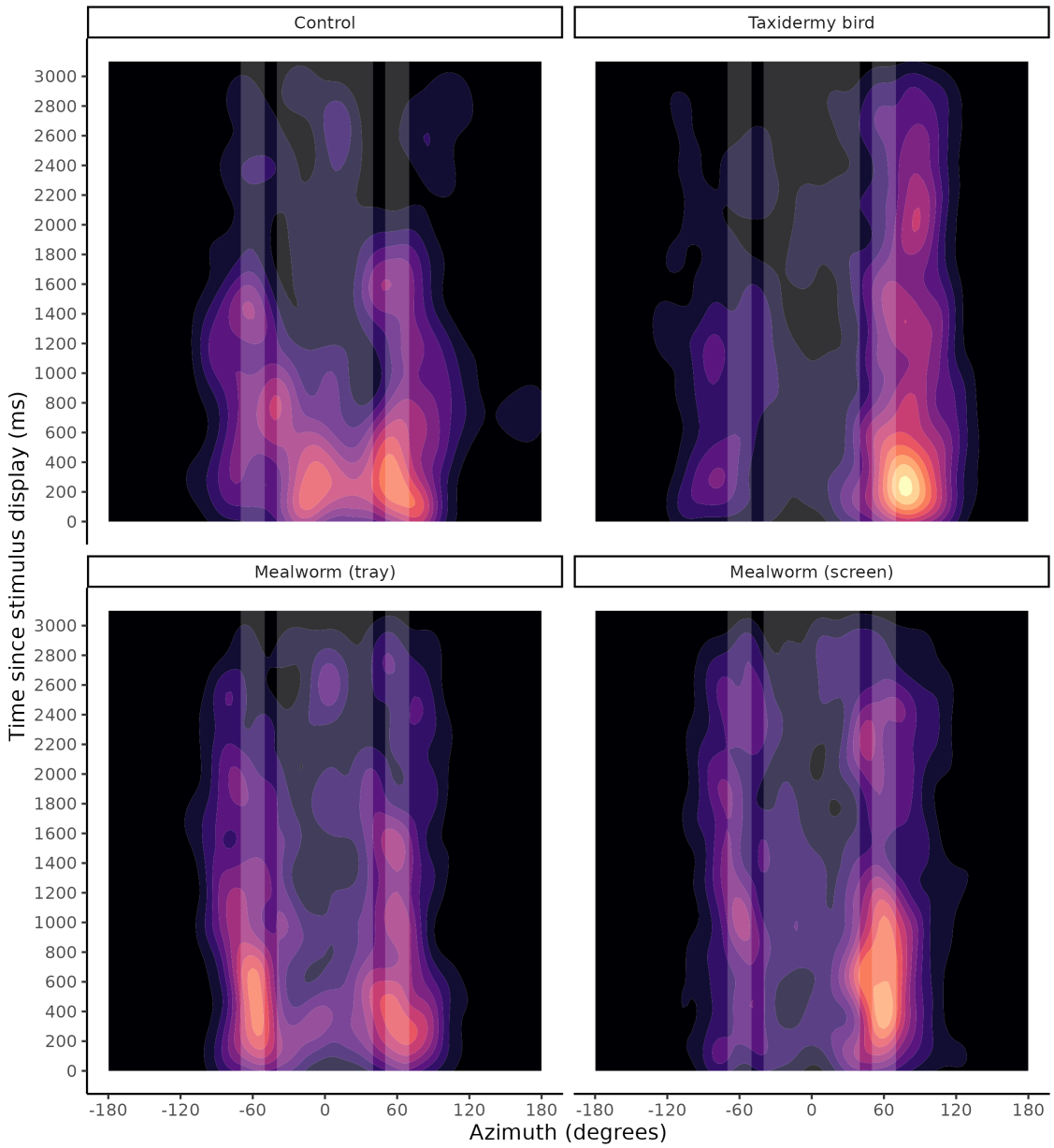

Figure 3: **Gaze over time.** Density of head azimuth over the 3s after the stimulus was displayed. Informed by our mixture model, data was subset to head elevations between -15 and 5 degrees to exclude probable non-looking observations. White rectangles indicate estimated optic axes from our mixture model, and likely binocular field based on estimates from related species (see discussion).

2.4 Population Monitoring

Next, we illustrate how 3D-SOCS can be used as a non-invasive population monitoring tool that provides insight into the distribution of weights of birds in a population without having to catch and weigh wild birds. We extracted the convex hull volume of each individual. The body volume of great tits and blue tits as estimated from 3D-SOCS significantly predicted body weight from manually weighing undertaken when the birds were ringed, while controlling for species (scaled volume $\hat{\beta}$ (95% HPDI) $= 0.3(0.15, 0.44)$, Table S8). The

contrast between varying intercepts for species also showed that the body volume can significantly distinguish between great tit and blue tits ($\hat{\alpha}_{GT} - \hat{\alpha}_{BT} = 2.03(1.59, 2.47)$). This reflects the known distribution of weights in these two species, where great tits are heavier than blue tits, with little to no overlap (mean weight GT=17.9g, BT=11.6g). Therefore, body volume can potentially be used as a proxy for body weight, allowing for the weight of individual birds to be estimated using 3D-SOCS, providing valuable information about the overall spatial and temporal variation in populations.

3 Discussion

We present 3D-SOCS as a flexible, scalable and inexpensive markerless posture tracking system for wild animals, and illustrated its capabilities with a common study system in behavioral ecology. To the best of our knowledge, this is the first system showcasing 3D markerless posture tracking of multiple animals from a wild population. Our evaluation of the system's accuracy demonstrated that 3D-SOCS has no consistent systematic error in estimating location and angles of the head, and can recover keypoint estimates within 3mm of their true location. Our experiment demonstrated the value of 3D tracking in examining great tits' visual field use in response to visual stimuli. By tracking the birds' head orientations, we were able to estimate the most frequently used visual field areas, primarily located near the optic axes, and also explored their lateralization of eye use. Finally, we show that body volume estimation can be a useful proxy for species differentiation and body weight. We provide detailed documentation (available in supplementary text S1.1 and S1.2), and code (available in Section 5) for other researchers to deploy the system in the field who wish to quantify fine-scale behaviours in wild birds.

We validated 3D-SOCS performance by conducting a series of system accuracy tests in a marker-based motion capture system. Our evaluation demonstrated that the keypoint measurements in 3D-SOCS are normally distributed, with no systematic deviation from their ground truth values, and that errors typically fall within a 3mm range. Our tests using only two cameras found error rates comparable to those with six cameras, albeit with a trade-off in frame loss. Generally, more cameras improve frame loss. Also importantly, 3D-SOCS is not limited to our setup with a metal cage. As long as two or more cameras can be mounted to provide multiple views of a scene, and a 2D posture detection model is available, 3D-SOCS can be effectively deployed across a variety of study systems and species.

We used 3D-SOCS to identify the visual field use of great tits with 3D tracking of head orientations. We found that the location of great tit's optic axes are approximately $\pm 60^\circ$ azimuth and -5° elevation. This result is largely consistent with previous literature in passerines. For example, captive zebra finches (*Taeniopygia guttata*) and crows (*Corvus macrorhynchos*) were previously found to have optic axes around $\pm 60^\circ$ [65, 70]. For future researchers who would like to use 3D-SOCS to quantify the lateral gaze of great tits, we would suggest an uncertainty cone of 10° around our estimate of optic axes to account for both eye movement and rotational error. It was difficult to quantify the precise boundaries of the binocular field from our data, although birds certainly used their binocular field especially when they approached food items (Figures 2, 3). A prior study found that the binocular fields of tufted titmice and chickadees fell between $\pm 26^\circ$ azimuth when eyes were not rotated forward, and between $\pm 40^\circ$ when eyes were rotated forward [72]. Given these species' all fall in family Paridae, we speculate that tits' values would be similar, although this would require empirical verification.

Furthermore, we observed lateralization in eye use in response to the different stimuli we presented. Hemispheric specialization could underlie the preferential use of either eye, although it has been argued that severe asymmetry should not be selectively favored [73]. Our results suggest that at the population level, tits were biased towards using their right eye (implicating the left hemispheric bias) especially when observing the taxidermy bird. In passerines, pigeons, and domestic chicks, the left hemisphere is generally superior in feature discrimination and memorization tasks [74]. It is therefore not surprising that great tits would also favor the right eye when attending to novel stimuli. However, this finding does contrast with the hypothesis that social cognition is dominated by the right hemisphere [75, 76], which is phylogenetically conserved [77–80]. However, quail (*Coturnix sp.*) were found to regard familiar individuals with their left eye, but to prefer the right eye for unknown conspecifics [81]. It may be that the tits follow a similar pattern, or that there was some uncertainty as to whether the stimuli represented a conspecific. This could be further explored by presenting a heterospecific stimulus.

Although our illustrative experiments have highlighted the capabilities of 3D-SOCS, it is important to note that this system can be utilized across a variety of experimental paradigms. 3D-SOCS has the most potential for wild studies, especially when paired with cognitive tasks or experimental presentations, such as our stimulus presentation. However, 3D-SOCS could also allow for more naturalistic behavior in laboratory experiments

that measure gaze, which traditionally require restricting movement of animals [82]. Moreover, whether used in the wild or the lab, the tracked 3D postures can be utilized to automatically generate ethograms using either unsupervised or supervised algorithms [41]. The classified behaviors could then be used to quantify fine-scale variation in various interactions or displays such as fights or courtship displays at lekking sites [50, 83]. Finally, our finding that estimated body size serves as a reliable proxy for weight demonstrates a practical use-case for 3D-SOCS in quantifying body condition. This allows for repeated measurements without the need for catching, with various potential applications including monitoring morphological changes over time, such as in the observed link between climate change and decreasing body mass in great tits [84].

In conclusion, we have demonstrated how 3D-SOCS offers a low-cost, non-invasive method for gaining unprecedented insights into the fine-scale detail of the behavior of wild passerine birds. Our case study on the visual field of great tits suggests the possibility of inferring the area centralis in birds. Finally, we emphasize that 3D-SOCS is versatile in many respects: it can be deployed remotely in the field, it requires as few as two cameras, and with the appropriate inference models it could potentially be applied to a wide range of species.

4 Material and Methods

4.1 Study species and site

Great tits (Parus major) are a common passerine songbird with a generalist diet found in woodlands and gardens throughout Europe and Asia [85]. They are an established study species in behavioral ecology [86], with extensive previous study of their breeding, foraging and social behaviour [7, 26, 87–90].

The study took place in the woodland surroundings of the Max Planck Institute of Animal Behavior, located in Möggingen, Germany (47.765088, 8.997779). Wild tits were caught using mist-nets within a 1 km radius of the institute. Once caught, birds were ringed and fitted with a passive transponder (PIT tag) under ethical approval from the Regierungspräsidium Freiburg (35-9185.81/G-22/070). While ringing, we also recorded age class (juvenile < 1 yr, adult ≥ 1 yr), sex based on plumage, weight, tarsus length and wing length.

4.2 3D-SOCS

We constructed a 1x1m metal cage in which we could position cameras, a food source, and a stimulus presentation device (Figure 1A). The cage mesh was large enough to let small passerine birds in and out, but protected objects inside of the cage from squirrels and other non-target visitors. We covered the left, right and bottom of the cage with white PVC plate to protect to prevent birds from entering from the side, and to prevent visual distractions from outside of the cage. All electronics required for the cage to operate were housed in a box which sat on top of the cage, protected by a roof fabricated from corrugated plastic and wood. The system was powered by a rechargeable battery unit.

Within the cage, we positioned six frame-synchronized cameras at various angles focused on a food source where birds were expected to land, consisting of a small bowl filled with a sunflower seed and peanut mixture (Figure 1B). Each camera was controlled by a RaspberryPi 4 Compute Module (CM4) micro-computer using PiCamera2 python library. Sub-millisecond frame synchronization was achieved by taking advantage of Precision Time Protocol (PTP), a Linux protocol which can synchronize the clocks of networked computers. A seventh CM4 was designated as the lead, and it acquired time from GPS satellites using a Timebeat GPS/GNSS module. The other six CM4 synchronized their hardware clocks to this time using Timebeat software implementation of PTP. A custom encoder in PiCamera2 disciplined the framerate of each camera to the synchronized system clock of its CM4, achieving frame captures on each camera with sub-millisecond differences. Recording was triggered by a computer vision based motion-detection algorithm, which also ran on the lead CM4. When motion was detected, it triggered the other six cameras to begin recording using a GPIO signal. Recordings lasted for 60 seconds each. Further details of the synchronized video capture are available in Text S1.1, and online in our github repository (see section 5).

4.3 Stimulus presentation device

In order to standardize the stimulus presentation to individuals, we constructed a small box which held a stimulus behind a piece of polymer dispersed liquid crystal (PDLC) film, a material that is in an opaque or transparent state depending on whether a current is supplied. The device was controlled by a RaspberryPi 4 microcomputer with custom Python software. The software monitored a perch equipped with an energized antenna that was placed directly over the food source. When a PIT tagged bird landed on the perch, an RFID

reader (PriorityOne brand) read the individual identity of the bird. After a short delay of 0.5s, the PDLC film then became transparent, revealing the stimulus. It remained transparent until the bird was detected as leaving the perch. The device recorded the bird's identity and a timestamp of when the bird arrived and left.

4.4 Experimental study design

We set up our 3D-SOCS cage in a small patch of trees that was connected to a larger forest. Prior to starting the experiment, we placed a feeder inside of the cage for one week to habituate birds to enter the cage. We began with a 7 day habituation phase, where we placed a large plate of seeds and mealworms in the arena to habituate birds to 3D-SOCS. We also used this period to troubleshoot the data collection pipeline. We then replaced the food plate with the stimulus presentation device, and recorded observations over 5 days (November 5th to November 10th, 2023). We tested birds in four conditions presenting four different stimuli: 1) No stimulus condition as control for the clicking sound of the switchable glass turning on, 2) food stimulus condition presenting a small dish filled with mealworms, a high-reward food item, 3) food video stimulus condition presenting a looped animation of a single mealworm that wiggled in place on a screen (Huawei MediaPad T2-8 Pro, 60 Hz 8" IPS screen with 1200x1920px resolution), and 4) social stimulus condition presenting a taxidermy female adult great tit. Using a taxidermy bird allowed us to control the identity, posture and location of the observed bird.

4.5 Datasets and model training

We trained two separate models, YOLOv8-large [67], an object detection model to localize a bird in frame, and DeepLabCut (DLC; [40]) with pre-trained ResNet50 as the 2D posture estimation model. For object detection, we manually annotated a dataset with 1601 frames randomly sampled across the views, with bounding boxes labelled "Great tit" and "Blue tit". Out of these images, we further annotated 905 bird instances with 14 unique keypoints if it was visible. These keypoints included: bill tip, bill base, L/R eye, L/R neck, L/R cheek, head back, L/R wing bar front, L/R wing bar back and tail base (Figure 1C). We used default augmentation parameters for YOLO and custom image augmentation and learning rate for DLC. The complete keypoint and bounding box dataset is available (see section 5).

4.6 3D posture estimation

We briefly describe the processing pipeline, although full details can be found in section S1.2. We used a laser-printed ChArUco checkerboard generated using the opencv library for calibration of camera intrinsics and extrinsics at the beginning of each day of data collection. Intrinsic calibration refers to each camera's internal parameters (e.g. focal length, lens distortion). Extrinsic calibration refers to parameters that describe the relative positions between each camera in the scene. Figure 1D illustrates the full tracking pipeline for 3D posture estimation. We first detected bounding boxes of all birds present in each frame using YOLOv8 and then tracked birds across frames in 2D using BoT-SORT to provide a unique identity for each bird [91]. Bounding boxes were used to crop frames of birds, which were then fed into a single animal DeepLabCut model [40], to generate 2D keypoint predictions for each bird per view. Next, we estimated correspondence between individuals across views for every frame through iterative matching, based on Huang et al. [92]. We used the matched correspondences to triangulate 2D detections into 3D posture estimates per frame using bundle adjustment by minimizing the reprojection error of a predicted 3D point across all cameras.

We implemented several post-processing steps, including interpolation and smoothing. We filtered outlier 3D keypoints when the reprojections were outside of any camera frame. We removed keypoints that fell beyond a set boundary of the head or body, defined by 1 standard deviation away from the mean position of the overall object. Finally, since head keypoints were rigid relative to each other for each bird, we calculated an average "skull" for each individual by taking the median measurements for each point in the head coordinate system. If at least 3 points of a bird's skull were detected, we could use the average skull to infer the position of any missing keypoints by computing its rotation and translation in space. We leveraged this when testing stereo cameras as a minimal set-up for 3D-SOCS. Stereo cameras are inherently more challenging since the number of points detected by both cameras is limited. However, as long as 3 points were detected, we used a species-specific median skeleton to fill in missing keypoints, providing 3D keypoint estimates of the whole bird.

4.7 System accuracy test

To validate the overall accuracy of 3D-SOCS, we deployed the system with the same camera configuration in a highly accurate marker-based motion capture system SMART-BARN [35]. The goal of the test was to determine the mean and variance of the absolute distance error and rotational error of a given object for 3D-SOCS, against estimates from SMART-BARN. To inform future researchers on the trade-off between camera settings, we also tested two camera resolutions: a narrow field of view (1920x1080px) that is focused on the food tray, and a wide field of view (1640x1280px) which can capture a wider area of the 1 cubic-meter platform.

We refer to section S1.4 for a detailed description of the workflow. In brief, a taxidermy great tit and an ArUco QR code were systematically rotated and moved around the volume for a total of around 20 minutes per camera resolution. We then measured the distance and rotation error of a set of head keypoint (bill tip, L/R eyes) by comparing the detected point with a reference QR code. Since the environment and taxidermy great tit looks slightly different from the data collected by 3D-SOCS in the field, we also fine-tuned the DeepLabCut and YOLO model using 100 additional annotated frames of the taxidermy tit, to provide the most ideal conditions to evaluate the accuracy of the system.

4.8 Field of view and lateralization analysis

Assuming that birds were most likely to regard the stimulus immediately after its display, we limited the posture data to include only the 3 seconds following the stimulus onset. For each frame, we transformed the 3D coordinate of the stimulus to the local head coordinate system, then calculated the position of the stimulus using a modified spherical coordinate system, where azimuth and elevation of 0 corresponded to the forward vector between the mid-point of the eyes and the bill tip (Figure S2).

We used a set of filters to subset frames where a bird was likely attending to the stimulus. Birds use saccadic head movements, characterized by a rapid movement of the head followed by a longer fixation [93].It should be noted that due to the limited temporal resolution of our RGB cameras (30 fps), our system was unable to detect small and brief head-saccades. Our aim was to extract head fixations that were separated by relatively large saccade-like movements, rather than to detect the head-saccades themselves. We first computed the angular change of a bird's head between each frame and defined saccades as changes of $> 500°$ per second (16.7°per frame). We then categorized frames between these detected saccades as fixation frames. Since the detected saccades that were very short or very long could potentially be attributed to noise in our 3D posture estimation, we disregarded those saccades that were shorter than two frames ($< 66ms$) or longer than five frames ($> 165ms$). We also disregarded saccade amplitudes of $< 33°$, manually extracted from Figure 2, and fixation periods of 1 frame or less. To exclude frames where birds were eating seeds with their heads down (likely focusing on the food), we filtered out frames in which the bill tip was within a defined cuboid around the food tray.

We fit a Bayesian multivariate mixture model to this data to estimate the optic axes. Each observation was a two-dimensional vector representing the head azimuth and elevation relative to the stimulus. The likelihood of an observation was modeled as a mixture of $K = 2$ two-dimensional Gaussians (representing the lateral fields of their left and right eyes) with each component's mean vector composed of an azimuth and elevation component ($\mu_{\text{az.},s}[k], \mu_{\text{elev.}}$) and scale vector ($\sigma_{\text{az.}}[k], \sigma_{\text{elev.}}$). Parameters and priors are summarized in table S9. Azimuth means, standard deviations, and mixing proportions were estimated for each stimulus, but the elevation was estimated via complete pooling.

As this model was intended to estimate the head elevation and optic axes and could not account for varying effects of individual ID, we fit a secondary Bayesian logistic model with varying effects. We subset data to measurements of head elevation between $-15°$ and $5°$ (informed by the first model and rotation error from the system accuracy tests) to conservatively only include observations of birds likely attending to stimuli. We coded a binary variable representing when a bird favored their right lateral field as 1 when azimuth was between $50°$ and $70°$, and 0 when azimuth was $-50°$ and $-70°$. This variable was modelled as a Bernoulli distribution with probability θ, composed of a sum of varying intercepts for stimulus type and individual ID.

4.9 Estimation of body volume as proxy for weight

Using body keypoints, we constructed a median body skeleton for each individual bird across all trials, then defined a sparse mesh using the trimesh python library [94]. From this, we computed a 3D convex hull volume (supplementary text S1.2) as a proxy for body size. To test the relationship between the convex hull and weight in great tits, we used a Bayesian linear model where scaled weight was modeled as a normal distribution whose mean was predicted by scaled volume and a varying intercept for species class. Each bird's weight

was calculated as the mean of all its measured weights across captures from our catalogue of ringing data, excluding its weight as a nestling.

4.10 Bayesian model fitting

All models described above were fit using Hamiltonian Markov Chain Monte Carlo (MCMC). Models were run using 5 chains and a minimum of 5000 iterations, with 2500 warm-up iterations, with all estimates based on over at least 1976 effective samples from the posterior (range: 1976–18955). All models were fitted using R v. 4.3.2 with Stan v. 2.32.2 [95, 96]. Good model convergence was confirmed from evaluating rank histograms and ensuring parameters had Gelman–Rubin's statistic ≤ 1.01 [97]. Given the complexity of the Gaussian mixture model used to estimate the optic axes, we further validated this model using simulated data (see section 5).

5 Data Availability

Code and data for reproducing statistical analyses and main text figures will be publicly available upon acceptance. Code and instructions for creating and running 3D-SOCS yourself are available at https://github.com/alexhang212/3D-SOCS. Sample datasets for running the code and bounding box and keypoint annotations can be downloaded here [98]: https://doi.org/10.17617/3.ZQMOJ3.

6 Acknowledgements

This work was supported by the Max Planck Society and the Centre for the Advanced Study of Collective Behaviour, funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy (EXC 2117-422037984). LMA and MC were also partly funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00056. We thank: Gustavo Alarcón-Nieto for helping with bird ringing; the members from the Comparative Sociality and Cognition Lab for the support and feedback; MPI-AB veterinarians Inge Mueller and Daniel Zuniga for their support.

7 Author Contributions

Conceptualisation - M.C., A.H.H.C., L.M.A., F.K.;
Methodology - M.C., A.H.H.C., L.M.A., F.K;
Software - M.C., A.H.H.C.;
Investigation - M.C., A.H.H.C.;
Resources - M.C., A.H.H.C., F.K.;
Writing Original Draft - M.C., A.H.H.C.;
Writing Review & Editing - M.C., A.H.H.C., L.M.A., F.K;
Visualization - M.C., A.H.H.C.;
Supervision - L.M.A., F.K.;
Funding Acquisition - M.C., L.M.A., F.K.;

8 Competing Interests

The authors declare no competing interests, financial or otherwise.

References

1. Farine, D. R., Aplin, L. M., Garroway, C. J., Mann, R. P. & Sheldon, B. C. Collective decision making and social interaction rules in mixed-species flocks of songbirds. *Animal behaviour* **95,** 173–182 (2014).
2. Farine, D. R. *et al.* The role of social and ecological processes in structuring animal populations: a case study from automated tracking of wild birds. *Royal Society Open Science* **2,** 150057 (2015).
3. Flack, A., Nagy, M., Fiedler, W., Couzin, I. D. & Wikelski, M. From local collective behavior to global migratory patterns in white storks. *Science* **360,** 911–914 (2018).

4. Flack, A. *et al.* New frontiers in bird migration research. *Current Biology* **32,** R1187–R1199 (2022).

5. Klarevas-Irby, J. A., Wikelski, M. & Farine, D. R. Efficient movement strategies mitigate the energetic cost of dispersal. *Ecology Letters* **24,** 1432–1442 (2021).

6. Farine, D. R., Aplin, L. M., Sheldon, B. C. & Hoppitt, W. Interspecific social networks promote information transmission in wild songbirds. *Proceedings of the Royal Society B: Biological Sciences* **282,** 20142804 (2015).

7. Aplin, L. M. *et al.* Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature* **518,** 538 (2015).

8. Fang, C., Huang, J., Cuan, K., Zhuang, X. & Zhang, T. Comparative study on poultry target tracking algorithms based on a deep regression network. *Biosystems Engineering* **190,** 176–183 (2020).

9. Hawkes, L. A., Broderick, A. C., Godfrey, M. H. & Godley, B. J. Climate change and marine turtles. *Endangered Species Research* **7,** 137–154 (2009).

10. Nourani, E. *et al.* Seabird morphology determines operational wind speeds, tolerable maxima, and responses to extremes. *Current Biology* **33,** 1179–1184 (2023).

11. Couzin, I. D. & Heins, C. Emerging technologies for behavioral research in changing environments. *Trends in Ecology & Evolution* **38,** 346–354 (2023).

12. Lahoz-Monfort, J. J. & Magrath, M. J. A comprehensive overview of technologies for species and habitat monitoring and conservation. *BioScience* **71,** 1038–1062 (2021).

13. Tuia, D. *et al.* Perspectives in machine learning for wildlife conservation. *Nature communications* **13,** 792 (2022).

14. Wilmers, C. C. *et al.* The golden age of bio-logging: How animal-borne sensors are advancing the frontiers of ecology. *Ecology* **96,** 1741–1753 (2015).

15. Wild, T. A. *et al.* Internet on animals: Wi-Fi-enabled devices provide a solution for big data transmission in biologging. *Methods in Ecology and Evolution* **14,** 87–102 (2023).

16. Bonter, D. N. & Bridge, E. S. Applications of radio frequency identification (RFID) in ornithological research: a review. *Journal of Field Ornithology* **82,** 1–10 (2011).

17. Iserbyt, A., Griffioen, M., Borremans, B., Eens, M. & Müller, W. How to quantify animal activity from radio-frequency identification (RFID) recordings. *Ecology and Evolution* **8,** 10166–10174 (2018).

18. König, B. *et al.* A system for automatic recording of social behavior in a free-living wild house mouse population. *Animal Biotelemetry* **3,** 1–15 (2015).

19. Raulo, A. *et al.* Social and environmental transmission spread different sets of gut microbes in wild mice. *Nature Ecology & Evolution,* 1–14 (2024).

20. Maldonado-Chaparro, A. A., Forstmeier, W. & Farine, D. R. Relationship quality underpins pair bond formation and subsequent reproductive performance. *Animal Behaviour* **182,** 43–58 (2021).

21. Individual personalities predict social behaviour in wild networks of great tits (Parus major). *Ecology letters* **16,** 1365–1372 (2013).

22. Beck, K. B., Farine, D. R. & Kempenaers, B. Winter associations predict social and extra-pair mating patterns in a wild songbird. *Proceedings of the Royal Society B* **287,** 20192606 (2020).

23. Jolles, J. W. Broad-scale applications of the Raspberry Pi: A review and guide for biologists. *Methods in Ecology and Evolution* **12,** 1562–1579 (2021).

24. Wild, S., Alarcón-Nieto, G., Chimento, M. & Aplin, L. M. Manipulating actions: A selective two-option device for cognitive experiments in wild animals. *Journal of Animal Ecology* **92,** 1509–1519 (2023).

25. Cauchoix, M. *et al.* The OpenFeeder: A flexible automated RFID feeder to measure interspecies and intraspecies differences in cognitive and behavioural performance in wild birds. *Methods in Ecology and Evolution* **13,** 1955–1961 (2022).

26. Chimento, M., Alarcón-Nieto, G. & Aplin, L. M. Population turnover facilitates cultural selection for efficiency in birds. *Current Biology* **31,** 2477–2483 (2021).

27. Kings, M. *et al.* Wild jackdaws can selectively adjust their social associations while preserving valuable long-term relationships. *Nature Communications* **14,** 5103 (2023).

28. Beck, K. B. *et al.* Experimental manipulation of population density in a wild bird alters social structure but not patch discovery rate. *Animal Behaviour* **209,** 95–120 (2024).

29. Luxem, K. *et al.* Open-source tools for behavioral video analysis: Setup, methods, and best practices. *Elife* **12,** e79305 (2023).

30. Mathis, M. W. & Mathis, A. Deep learning tools for the measurement of animal behavior in neuroscience. *Current opinion in neurobiology* **60,** 1–11 (2020).

31. Mathis, A., Schneider, S., Lauer, J. & Mathis, M. W. A primer on motion capture with deep learning: principles, pitfalls, and perspectives. *Neuron* **108,** 44–65 (2020).

32. Ehlman, S. M. *et al.* Leveraging big data to uncover the eco-evolutionary factors shaping behavioural development. *Proceedings of the Royal Society B* **290,** 20222115 (2023).

33. Griebling, H. J. *et al.* How technology can advance the study of animal cognition in the wild. *Current Opinion in Behavioral Sciences* **45,** 101120 (2022).

34. Walter, T. & Couzin, I. D. TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *Elife* **10,** e64000 (2021).

35. Nagy, M. *et al.* SMART-BARN: Scalable multimodal arena for real-time tracking behavior of animals in large numbers. *Science Advances* **9,** eadf8068 (2023).

36. Alarcón-Nieto, G. *et al.* An automated barcode tracking system for behavioural studies in birds. *Methods in Ecology and Evolution* **9,** 1536–1547 (2018).

37. Waldmann, U. *et al.* 3D-MuPPET: 3D Multi-Pigeon Pose Estimation and Tracking. *International Journal of Computer Vision.* ISSN: 1573-1405. https://doi.org/10.1007/s11263-024-02074-y (May 2024).

38. Naik, H. *et al. 3D-POP-An automated annotation approach to facilitate markerless 2D-3D tracking of freely moving birds with marker-based motion capture* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), 21274–21284.

39. Chiara, V. & Kim, S.-Y. AnimalTA: A highly flexible and easy-to-use program for tracking and analysing animal movement in different environments. *Methods in Ecology and Evolution* (2023).

40. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience* **21,** 1281–1289 (2018).

41. Dunn, T. W. *et al.* Geometric deep learning enables 3D kinematic profiling across species and environments. *Nature methods* **18,** 564–573 (2021).

42. Bala, P. C. *et al.* Openmonkeystudio: Automated markerless pose estimation in freely moving macaques. *BioRxiv,* 2020–01 (2020).

43. Romero-Ferrero, F., Bergomi, M. G., Hinz, R. C., Heras, F. J. & De Polavieja, G. G. Idtracker. ai: tracking all individuals in small or large collectives of unmarked animals. *Nature methods* **16,** 179–182 (2019).

44. Gosztolai, A. *et al.* LiftPose3D, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nature methods* **18,** 975–981 (2021).

45. Karashchuk, P. *et al.* Anipose: a toolkit for robust markerless 3D pose estimation. *Cell reports* **36** (2021).

46. Pereira, T. D. *et al.* SLEAP: A deep learning system for multi-animal pose tracking. *Nature methods* **19,** 486–495 (2022).

47. Koger, B. *et al.* Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. *Journal of Animal Ecology* (2023).

48. Graving, J. M. *et al.* DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* **8,** e47994 (2019).

49. Schad, L. & Fischer, J. Opportunities and risks in the use of drones for studying animal behaviour. *Methods in Ecology and Evolution* **14,** 1864–1872 (2023).

50. Janisch, J. *et al.* Video recording and analysis of avian movements and behavior: insights from courtship case studies. *Integrative and Comparative Biology* **61,** 1378–1393 (2021).

51. Chan, A. H. H., Liu, J., Burke, T., Pearse, W. D. & Schroeder, J. Comparison of manual, machine learning, and hybrid methods for video annotation to extract parental care data. *Journal of Avian Biology,* e03167 (2023).

52. Dunkley, K., Dunkley, A., Drewnicki, J., Keith, I. & Herbert-Read, J. E. A low-cost, long-running, open-source stereo camera for tracking aquatic species and their behaviours. *Methods in Ecology and Evolution* (2023).

53. Ferreira, A. C. *et al.* Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution* **11,** 1072–1085 (2020).

54. Schofield, D. P. *et al.* Automated face recognition using deep neural networks produces robust primate social networks and sociality measures. *Methods in Ecology and Evolution* **14,** 1937–1951 (2023).

55. Jiang, L., Lee, C., Teotia, D. & Ostadabbas, S. Animal pose estimation: A closer look at the state-of-the-art, existing gaps and opportunities. *Computer Vision and Image Understanding,* 103483 (2022).

56. Lauer, J. *et al.* Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods* **19,** 496–504 (2022).

57. An, L. *et al.* Three-dimensional surface motion capture of multiple freely moving pigs using MAMMAL. *Nature Communications* **14,** 7727 (2023).

58. Han, Y. *et al.* Multi-animal 3D social pose estimation, identification and behaviour embedding with a few-shot learning framework. *Nature Machine Intelligence,* 1–14 (2024).

59. Joska, D. *et al. AcinoSet: a 3D pose estimation dataset and baseline models for Cheetahs in the wild* in *2021 IEEE international conference on robotics and automation (ICRA)* (2021), 13901–13908.

60. Xiao, S. *et al.* Multi-view Tracking, Re-ID, and Social Network Analysis of a Flock of Visually Similar Birds in an Outdoor Aviary. *International Journal of Computer Vision* **131,** 1532–1549 (2023).

61. Anderson, D. J. & Perona, P. Toward a science of computational ethology. *Neuron* **84,** 18–31 (2014).

62. Delacoux, M. & Kano, F. Fine-scale tracking reveals visual field use for predator detection and escape in collective foraging of pigeon flocks. *bioRxiv,* 2024–02 (2024).

63. Mitoyen, C., Quigley, C., Boehly, T. & Fusani, L. Female behaviour is differentially associated with specific components of multimodal courtship in ring doves. *Animal Behaviour* **173,** 21–39 (2021).

64. Kano, F., Naik, H., Keskin, G., Couzin, I. D. & Nagy, M. Head-tracking of freely-behaving pigeons in a motion-capture system reveals the selective use of visual field regions. *Scientific Reports* **12,** 19113 (2022).

65. Itahara, A. & Kano, F. Gaze tracking of large-billed crows (Corvus macrorhynchos) in a motion capture system. *Journal of Experimental Biology,* jeb–246514 (2024).

66. Itahara, A. & Kano, F. " Corvid Tracking Studio": A custom-built motion capture system to track head movements of corvids. *Japanese Journal of Animal Psychology* **72,** 1–16 (2022).

67. Jocher, G., Chaurasia, A. & Qiu, J. *Ultralytics YOLO* version 8.0.0. Jan. 2023. https://github.com/ultralytics/ultralytics.

68. Kane, G. A., Lopes, G., Saunders, J. L., Mathis, A. & Mathis, M. W. Real-time, low-latency closed-loop feedback using markerless posture tracking. *Elife* **9,** e61909 (2020).

69. Butler, S. R., Templeton, J. J. & Fernández-Juricic, E. How do birds look at their world? A novel avian visual fixation strategy. *Behavioral ecology and sociobiology* **72,** 1–11 (2018).

70. Bischof, H.-J. The visual field and visually guided behavior in the zebra finch (Taeniopygia guttata). *Journal of Comparative Physiology A* **163,** 329–337 (1988).

71. Pratt, D. W. Saccadic eye movements are coordinated with head movements in walking chickens. *Journal of Experimental Biology* **97,** 217–223 (1982).

72. Moore, B. A., Doppler, M., Young, J. E. & Fernández-Juricic, E. Interspecific differences in the visual system and scanning behavior of three forest passerines that form heterospecific flocks. *Journal of Comparative Physiology A* **199,** 263–277 (2013).

73. Vallortigara, G., Cozzutti, C., Tommasi, L. & Rogers, L. J. How birds use their eyes: opposite left-right specialization for the lateral and frontal visual hemifield in the domestic chick. *Current Biology* **11,** 29–33 (2001).

74. Prior, H. Lateralization of spatial orientation in birds. *Behavioural and morphological asymmetries in vertebrates,* 75–85 (2006).

75. Daisley, J. N., Mascalzoni, E., Rosa-Salva, O., Rugani, R. & Regolin, L. Lateralization of social cognition in the domestic chicken (Gallus gallus). *Philosophical Transactions of the Royal Society B: Biological Sciences* **364,** 965–981 (2009).

76. Salva, O. R., Daisley, J. N., Regolin, L. & Vallortigara, G. Lateralization of social learning in the domestic chick, Gallus gallus domesticus: learning to avoid. *Animal Behaviour* **78,** 847–856 (2009).

77. De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. & Fazio, F. Prosopagnosia can be associated with damage confined to the right hemisphere—an MRI and PET study and a review of the literature. *Neuropsychologia* **32,** 893–902 (1994).

78. Hauser, M. D. Right hemisphere dominance for the production of facial expression in monkeys. *Science* **261,** 475–477 (1993).

79. Kendrick, K. M. Brain asymmetries for face recognition and emotion control in sheep. *Cortex* **42,** 96–98 (2006).

80. Sovrano, V. A., Rainoldi, C., Bisazza, A. & Vallortigara, G. Roots of brain specializations: preferential left-eye use during mirror-image inspection in six species of teleost fish. *Behavioural brain research* **106,** 175–180 (1999).

81. Zucca, P. & Sovrano, V. A. Animal lateralization and social recognition: quails use their left visual hemifield when approaching a companion and their right visual hemifield when approaching a stranger. *Cortex* **44,** 13–20 (2008).

82. Tyrrell, L. P., Butler, S. R., Yorzinski, J. L. & Fernández-Juricic, E. A novel system for bi-ocular eye-tracking in vertebrates with laterally placed eyes. *Methods in Ecology and Evolution* **5,** 1070–1077 (2014).

83. Quesada, J., Guallar, S., Navalpotro, H., Carrillo-Ortiz, J. G. & Senar, J. C. Recognizing interspecific dominance signals? Blue tits adjust nest defence based on great tit's black bib size. *Ethology,* e13460 (2024).

84. Furness, E. N. & Robinson, R. A. Long-term declines in winter body mass of tits throughout Britain and Ireland correlate with climate change. *Ecology and Evolution* **9,** 1202–1210 (2019).

85. Perrins, C. M. *British tits* (HarperCollins, 1979).

86. Gibb, J. The breeding biology of the Great and Blue titmice. *Ibis* **92,** 507–539 (1950).

87. Fisher, J. & Hinde, R. A. The opening of milk bottles by birds. *British Birds* **42,** 347–357 (1949).

88. Lefebvre, L. The opening of milk bottles by birds: evidence for accelerating learning rates, but against the wave-of-advance model of cultural transmission. *Behavioural Processes* **34,** 43–53 (1995).

89. Estók, P., Zsebők, S. & Siemers, B. M. Great tits search for, capture, kill and eat hibernating bats. *Biology letters* **6,** 59–62 (2010).

90. Morand-Ferron, J. & Quinn, J. L. Larger groups of passerines are more efficient problem solvers in the wild. *Proceedings of the National Academy of Sciences* **108,** 15898–15903 (2011).

91. Aharon, N., Orfaig, R. & Bobrovsky, B.-Z. BoT-SORT: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651* (2022).

92. Huang, C. *et al. End-to-end dynamic matching network for multi-view multi-person 3d pose estimation* in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII 16* (2020), 477–493.

93. Kano, F., Walker, J., Sasaki, T. & Biro, D. Head-mounted sensors reveal visual attention of free-flying homing pigeons. *Journal of Experimental Biology* **221** (2018).

94. Dawson-Haggerty et al. *trimesh* version 3.2.0. 2019. https://trimesh.org/.

95. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2018). https://www.R-project.org/.

96. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual* Version 2.27. 2021. http://mc-stan.org/.

97. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC (with Discussion). *Bayesian analysis* **16,** 667–718 (2021).

98. Chimento, M., Chan, H. H., Aplin, L. & Kano, F. *Additional dataset for: Peering into the world of wild passerines with 3D-SOCS: synchronized video capture for posture estimation* version V1. 2024. https://doi.org/10.17617/3.ZQMOJ3.