



This is a repository copy of *Comparison of manual, machine learning, and hybrid methods for video annotation to extract parental care data.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/206736/>

Version: Published Version

Article:

Chan, A.H.H. orcid.org/0000-0002-5405-7155, Liu, J., Burke, T. orcid.org/0000-0003-3848-1244 et al. (2 more authors) (2024) Comparison of manual, machine learning, and hybrid methods for video annotation to extract parental care data. *Journal of Avian Biology*, 2024 (3-4). e03167. ISSN 0908-8857

<https://doi.org/10.1111/jav.03167>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

JOURNAL OF AVIAN BIOLOGY

Short communication

Comparison of manual, machine learning, and hybrid methods for video annotation to extract parental care data

Alex Hoi Hang Chan^{1,2,3}, Jingqi Liu¹, Terry Burke⁴, William D. Pearse^{1,*} and Julia Schroeder^{1,*}

¹Department of Life Sciences, Imperial College London, Ascot, UK

²Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany

³Max Planck Institute of Animal Behaviour, Radolfzell, Germany

⁴Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

Correspondence: Alex Hoi Hang Chan (thealex.chan2@gmail.com)

Journal of Avian Biology

2023: e03167

doi: [10.1111/jav.03167](https://doi.org/10.1111/jav.03167)

Subject Editor: Wesley Hochachka

Editor-in-Chief: Staffan Bensch

Accepted 16 October 2023



Measuring parental care behaviour in the wild is central to the study of animal ecology and evolution, but it is often labour- and time-intensive. Efficient open-source tools have recently emerged that allow animal behaviour to be quantified from videos using machine learning and computer vision techniques, but there is limited appraisal of how these tools perform compared to traditional methods. To gain insight into how different methods perform in extracting data from videos taken in the field, we compared estimates of the parental provisioning rate of wild house sparrows *Passer domesticus* from video recordings. We compared four methods: manual annotation by experts, crowd-sourcing, automatic detection based on the open-source software DeepMeerkat, and a hybrid annotation method. We found that the data collected by the automatic method correlated with expert annotation ($r = 0.62$) and further show that these data are biologically meaningful as they predict brood survival. However, the automatic method produced largely biased estimates due to the detection of non-visitation events, while the crowd-sourcing and hybrid annotation produced estimates that are equivalent to expert annotation. The hybrid annotation method takes approximately 20% of annotation time compared to manual annotation, making it a more cost-effective way to collect data from videos. We provide a successful case study of how different approaches can be adopted and evaluated with a pre-existing dataset, to make informed decisions on the best way to process video datasets. If pre-existing frameworks produce biased estimates, we encourage researchers to adopt a hybrid approach of first using machine learning frameworks to preprocess videos, and then to do manual annotation to save annotation time. As open-source machine learning tools are becoming more accessible, we encourage biologists to make use of these tools to cut annotation time but still get equally accurate results without the need to develop novel algorithms from scratch.

Keywords: computer vision, DeepMeerkat, house sparrow, open-source, parental care



www.avianbiology.org

© 2023 The Authors. Journal of Avian Biology published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

*Contributed equally

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

Parental care behaviour is a life history trait that is commonly studied in a wide range of animals (Royle et al. 2012). Parental care is defined as any behaviour that increases the fitness of offspring (Clutton-Brock 1991, Royle et al. 2012) at a cost of the survival probability of parents (Trivers 1972), presenting a life history trade-off (Stearns 1992). While there are many forms of parental care (e.g. nest building, predator defence, incubation, and feeding; Royle et al. 2012), the feeding of young is considered particularly costly for parents because of the immense time and energy investments (Winkler and Wilkinson 1988, Owens and Bennett 1994). A large body of literature with a focus on birds, where 90% of species engage in parental care (Cockburn 2006), describes how nest visit frequency to dependent young is associated with aspects of an animal's life history. For example, studies of the life history trade-off between parent and offspring fitness (Schroeder et al. 2013), parental coordination (Wojczulanis-Jakubas et al. 2018, Ihle et al. 2019), parent-offspring conflict (Estramil et al. 2013), or ageing (Wilcoxon et al. 2010) have all used the frequency of parental visits to nests (provisioning rate) as a proxy of parental investment.

However, measuring provisioning rates in wild birds is labour-intensive and time-consuming. Data are traditionally collected by direct observation (Dunn and Cockburn 1996), which can be invasive by disturbing animals in the vicinity of their nest (Rose 2009). Less invasive methods include video recording (Nakagawa et al. 2007, García-Navas and Sanz 2010), radio tracking (Mitchell et al. 2012), and the use of radio-frequency tags and antennas at the nest (RFID; Ringsby et al. 2009, Mariette et al. 2011, Sánchez-Tójar et al. 2017). While radio tags allow visitation rates to be quantified over long periods of time, the technology is prone to missed detections (up to 20%; Mariette et al. 2011). Video analysis is more flexible and allows for other behaviours to be quantified, such as nest defence, copulations, or feeding load (Lendvai et al. 2015), but manual annotation of video data is time-consuming (Tuytens et al. 2014). Another alternative is crowd-sourcing (Desell et al. 2015), where students or citizen scientists are recruited to collect data. Crowd-sourcing is efficient in collecting data, while also being educational for students (Voss and Cooper 2010, Unger 2022), and increases engagement with the public. However, time has to be invested in training volunteers and in designing suitable software for citizen scientists (Desell et al. 2015, Root-Gutteridge et al. 2021). As the amount of video data collected by ecologists continues to increase, there is a need for more effective ways to extract data from video recordings (Weinstein 2018a).

Recent advances in deep learning (Borowiec et al. 2021) and computer vision (Weinstein 2018a) provide a solution to this problem, by allowing quick and consistent extraction of information from field data (Valletta et al. 2017). For example, machine learning methods have been successfully applied to solve problems with species identification (Wäldchen and Mäder 2018), bird song quantification (Pearse et al. 2018, Priyadarshani et al. 2018), social

behaviour measurement (Robie et al. 2017), and individual identification (Körschens et al. 2018, Bogucki et al. 2019, Schofield et al. 2019, Ferreira et al. 2020). Since computing resources are cheaper than human labour, such approaches have the potential to reduce the financial and time costs of data collection, evidenced by a recent increase in popularity for ecological applications (Borowiec et al. 2021, Tuia et al. 2022).

Recently, several open-source machine learning software packages have been developed to make methods accessible to biologists. Examples include software for tracking animals and behaviours in captive settings (Harmer and Thomas 2019, Pennington et al. 2019, Sridhar et al. 2019, Walter and Couzin 2021), pose estimation (Graving et al. 2019, Pereira et al. 2019, Lauer et al. 2022), or analysing timelapse videos (Weinstein 2015, 2018b). However, few studies have investigated whether these tools can be readily applied to pre-existing datasets, especially in field studies and long-term datasets. While much open-source software can be useful for processing videos collected in the field, biologists often do not know the appropriate software to use for their specific project, or are unsure whether investing time in constructing novel software for their study system will be worth the effort.

Here, we field tested alternative data collection methods – manual, crowd-sourcing, automatic (machine learning), and a manual/automatic hybrid method – in a model system with annotated parental care videos of wild house sparrows *Passer domesticus* on Lundy Island, UK (Nakagawa et al. 2007). We used DeepMeerkat (Weinstein 2018b), a popular machine learning-based framework that uses convolutional neural networks (CNNs) to detect movement events from wildlife monitoring videos. Despite the name, the software was initially designed for use in a hummingbird population (Weinstein 2018b, Marcot et al. 2019), and has been adapted for use in marine (Sheehan et al. 2020) and insect (Pegoraro et al. 2020, Mertens et al. 2021) systems. To the best of our knowledge, DeepMeerkat and its predecessor MotionMeerkat (Weinstein 2015) are the only open-source software tools designed specifically for counting occurrences of animals in ecological timelapse videos (Weinstein 2018a), hence we chose DeepMeerkat since it best fit the nature of our dataset. While Marcot et al. (2019) did a comparison of the efficacy of MotionMeerkat (Weinstein 2015) in the hummingbird system, there has been limited appraisal of the accuracy and effort trade-offs that researchers need to consider when applying these novel tools in their own systems. For this reason we set out to compare four data collection methods for measuring parental care visitation rates: 1) manual, 2) crowd-sourcing, 3) automatic (using machine learning), and 4) a hybrid of manual and automatic. Using a large multi-year dataset, we determined how well the tool performed in extracting parental care data, then compared the accuracy and time investment trade-off for each data collection method, to provide insight into how open-source tools can be readily incorporated into pre-established workflows to aid data collection.

Material and methods

We first describe the study system and the parental care video dataset used, then introduce a framework showing how a researcher can approach the use of machine learning tools with similarly large datasets. We then describe each of the data collection methods in detail, followed by evaluation metrics used to compare the methods. We then test a simple biological hypothesis of effort–fitness trade-offs, and finally quantitatively compare the accuracies of the four methods.

Study system

Data were collected from a population of house sparrows *P. domesticus* on Lundy Island (51°10'N, 4°40'W) in the Bristol Channel, UK. This population is part of a long-term study and has been monitored systematically since 2000, with > 99% of individuals each marked with a unique combination of colour rings, a metal ring with a unique number from the British Trust for Ornithology (Cleasby et al. 2011), and a unique passive-integrated transponder (Schroeder et al. 2011). Since house sparrows rarely fly over large bodies of water (Magnussen and Jensen 2017) very little immigration or emigration has taken place in the population (Schroeder et al. 2015). As a result, the population has high recapture rates with no trapping bias (Simons et al. 2015), and we have reliable life history data available for every individual (Schroeder et al. 2015).

Parental care data

The Lundy sparrow population is situated within an area of 0.2 km² around a small village, the only viable habitat on the island (Schroeder et al. 2011). Nest boxes for the sparrows are checked systematically to detect all breeding attempts throughout the summer breeding season (Cleasby et al. 2011). After eggs are found and the identities of the parents confirmed from their colour–ring combinations, 90-min videos (720 × 394, 25 fps) are recorded on Day 7 and Day 11 after egg hatching, using video cameras placed 2–5 m away from the nest box and with a field of view of at least 30 cm radius around the nest entrance to measure parental visitations (Nakagawa et al. 2007a for detailed procedure).

Data collection framework

Here, we describe a framework for formulating the approaches that researchers can use when incorporating machine learning tools for video analysis (Table 1). We categorised data collection methods into four types: 1) manual:

the traditional approach of data collection by expert researchers watching videos and transcribing behaviours manually, 2) crowd-sourcing: data collection manually by a large group of people such as students or the general public, 3) automatic: data collection by a machine learning tool, without human intervention, and 4) hybrid: a combination of both manual and automatic. Since the parental provisioning videos were taken in the field, there is no reliable ground truth for visitation rates because of differing video conditions, observer bias, or subjectivity of behaviours (Tuytens et al. 2014). For this reason, when comparing among methods, we are limited to assuming that manual annotation by experts is a suitable baseline. Finally, we also distinguish two main types of data that can be extracted with these methods: 1) bird presence, where birds are present within the video frame, and 2) bird visitation, a subset of the bird presence data, where birds enter the nest box or feed their young from the outside.

Method 1: manual

Manual annotation is based on an established protocol for parental care data for the study system (Nakagawa et al. 2007). During video annotation, birds were considered as feeding their young when entering and exiting the nest box and, in rare cases, when feeding behaviour could be observed through the nest box entrance with the parents not fully entering the nest box. We recorded when birds were perched outside the nest box as a separate behaviour that did not count towards visitation rate estimates. Visitation rates were calculated using the time period from the first visit of either parent until the end of the video, or until 90 min had elapsed from the first visit, whichever came first. We started counting from the first visit, and not from the beginning of the video, to allow time for habituation. The resulting time during which visits were scored was termed the effective observation time (Nakagawa et al. 2007a). The total number of visits by both parents was divided by the effective observational time to obtain the visitation rate (visits/h) as a measure of parental provisioning.

Between 2004 and 2015, videos (n = 2112) were manually annotated by graduate students and researchers working on the Lundy house sparrow project to obtain visitation rates. The dataset has contributed to multiple publications on the evolution of parental care (Nakagawa et al. 2007, Ihle et al. 2019, Schroeder et al. 2012, 2013, 2016, 2019), and we considered these data to be the manual expert (Table 1) dataset.

Method 2: crowd-sourcing

We also used a crowd-sourcing method where videos (n = 18) were provided to a cohort of 36 second-year undergraduate

Table 1. Summary of methods employed and definitions. Methods categorized into four types: (1) manual, (2) crowd-sourcing, (3) automatic, and (4) hybrid. Expert manual annotation used as a baseline to compare the other methods.

| Method | Definition | Sample size (number of videos) |
|--------------------|--|--------------------------------|
| (1) Manual | Annotation by expert researchers | 2112 |
| (2) Crowd-sourcing | Annotation by undergraduate students | 18 |
| (3) Automatic | Processing using machine learning pipeline without human intervention | 2629 |
| (4) Hybrid | Annotation of clips extracted from automatic pipeline by expert researcher | 18 |

students, who annotated full-length videos as part of their study. Students were first briefed on the annotation protocol as above. Each student was then assigned two videos to annotate. Since students are not trained experts, more than one student was assigned to every video to ensure no bias was introduced by individual observers. Finally, the average visitation rates for each video were calculated across observers. Each student also measured the approximate time it took for them to fully annotate each video.

Method 3: automatic

We processed videos collected between 2011 and 2019 ($n=2629$) with the open-source software DeepMeerkat (Weinstein 2018b), which detects image changes between frames and identifies moving objects in wildlife monitoring videos. While we considered other approaches such as object detectors (e.g. YOLOv8; Jocher et al. 2023) or background subtraction from the OpenCV library (Bradski 2000), we decided DeepMeerkat was the most appropriate since it is designed for wildlife videos, has a user-friendly interface, and has models pre-trained on similar tasks (Weinstein 2018b). Since DeepMeerkat only detects motion, events that are detected may not always accurately correspond to visitations. This method processes the data without human intervention, and has a 1:1 processing time on a single CPU (e.g. a 1.5 h video takes approximately 1.5 h; Marcot et al. 2019).

We first processed each video with DeepMeerkat, then merged movement events detected fewer than 40 frames apart (25 fps videos, i.e. 1.6 s) and more than two frames in length as the same bird presence event. We then further grouped the events into 7 s video clips, to allow downstream annotation (Method 4 below). We observed that most bird visitations can be captured well within the 7 s threshold. We tallied the number of events for each video, then divided the tally by the effective observation time (above) to obtain the automatic presence rate (in events/h; Fig. 1). Since most of the visitation events on Lundy sparrows were by parents entering the nest box, we broadly assumed each visitation event corresponded to two presence events (entering and exiting the nest box). So, the automatic presence rate was divided by two. Finally, since certain videos produced over-inflated presence rate measures due to the filming environment (e.g. camera shaking, continuous background movement), we removed any videos with an automatic presence rate of 72.7 events/h or more (1.31%

videos removed), since that was the maximum visitation rate we have ever recorded through expert manual annotation.

Method 4: hybrid

In the hybrid method, the same videos used in the crowdsourcing method ($n=18$) were first processed by the automatic method (Method 3) and then the video clips were manually annotated by experts to obtain visitation rates. This method filtered out events where birds were present, but not attending the nest, so manually converting the presence rate obtained from automatic processing (Method 3) to visitation rates. The time it took to annotate each video was recorded for comparison with other methods.

Comparing manual and automatic methods

To examine the reliability of the methods, we first compared the data obtained by manual annotation by experts with the automatic method (manual, $n=2112$; automatic, $n=2629$; Table 1). We used a Pearson's correlation test to determine whether both measures were correlated on videos where both methods were used ($n=781$), then used both measures in a case study to test the hypothesis that parental care was beneficial to offspring and parental fitness (Trivers 1972). We tested whether broods whose parents had higher presence or visitation rates had a higher fledging and recruitment success, with 'fledging' referring to a sparrow chick being successfully fledged from the nest, and recruits being fledglings that produced at least one genetic offspring in their lifetime. While we expected visitation rates by expert manual annotation to predict higher fitness from previous studies (Schroeder et al. 2013), we were also able to determine whether bird presence rates collected by automatic processing could similarly be a proxy for parental care behaviour.

We fitted two generalised linear mixed models with the number of fledglings and the number of recruits for each brood as respective response variables, against the visitation rates and presence rates as explanatory variables, using a Poisson link function. We then z-transformed both measures of parental provisioning to allow effect sizes between the data collected by manual and automatic methods to be compared. We used absolute count of fledglings and recruits as a proxy of fitness, because we were interested in how the absolute differences in provisioning rates in the brood level will

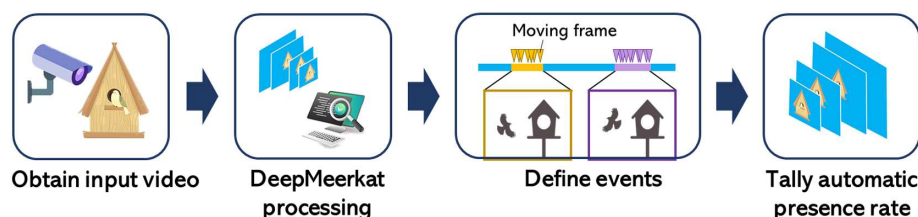


Figure 1. Automatic pipeline for processing parental provisioning videos of house sparrows on Lundy Island, UK. Videos were processed using open-source software DeepMeerkat (Weinstein 2018b), outputting frames with movement detected. Events were then defined by grouping movement frames that were less than 40 frames apart and at least two frames long. Finally, events for each provisioning video were tallied to obtain a presence rate estimate.

affect the fitness outcomes of the whole brood. Moreover, there is no way of knowing which offspring the parents are feeding within a brood, so we assumed that parental investment scales proportionately to the number of offspring (Schroeder et al. 2013). The clutch size of Lundy sparrows also has limited variation (4.2 ± 0.8 ; Westneat et al. 2014). To account for other effects that might affect fitness and parental care behaviour, we ran the models with the following specifications. Since videos were collected on Day 7 and Day 11 after hatching, we averaged the values collected on both days to match with the response variable, because there was only a single fitness estimate for each brood. However, since visitation rates increase with brood age (Schroeder et al. 2019), we also ran the models separately using the rates on Day 7 and Day 11 to ensure results were consistent. We also added the ages of the mother and father (Wiebe 2018), and hatch date (days after 1 April) as fixed effects to control for their effects on fitness outcomes. Since breeding success usually correlates with peak food abundance (Lack 1968, Cresswell and McCleery 2003), we added a quadratic fixed-effect term for hatch date, assuming peak food abundance in the middle of the breeding season. Next, the population has undergone routine cross-fostering, which is associated with slightly increased chick survival (Winney et al. 2015), hence we added a fixed factor for fostered status (yes/no) in all models. Note that the parental provisioning rates we used were always from the parents that actually visited, who were not always the genetic parents of the young (Lattore et al. 2019). We added the social parent IDs and the year as random effects to control for environmental effects (Rose et al. 1998), and repeatable visitation rates by individual parents (Nakagawa et al. 2007). Finally, the location of the nest box was added as a random effect to control for environmental effects (Schroeder et al. 2012).

We ran all models using the R package 'MCMCglmm' (Hadfield 2010) in R ver. 3.6.1 (www.r-project.org). The posterior distributions and autocorrelations were checked following Hadfield (2014) to ensure all fixed and random effects converged without violating any model assumptions. We defined a parameter estimate as statistically significant if the 95% credible interval did not overlap with zero.

Comparing annotation methods

Next, we quantitatively compared crowd-sourcing, automatic, and hybrid approaches for collecting data on parental care videos using expert manual annotation as a baseline. We standardised the measures by dividing the number of events detected (presence events for automatic, and visit events for crowd-sourcing and hybrid) by each method over the baseline, to obtain a proportion of events detected compared to the baseline. Since a proportion of 1 would show that a method detected the same number of events as the baseline, we carried out one-sample *t*-tests to determine whether each method was significantly different from the baseline (by setting the theoretical mean (μ) to 1). We then compared all methods with each other, using pairwise *t*-tests.

Finally, to assess the performance and sources of error of the automatic data collection method, we further analysed the data collected by hybrid expert annotation on video clips (Method 4; $n = 1650$ bird presence events over 18 videos). We first computed false positive rates by comparing the automatic results with what the expert annotators recorded. We calculated the proportion of events where bird visitation was falsely detected, both in the case where the bird was not present within the frame, or when the bird was present but not visiting the nest. We then computed false negative and true positive rates by calculating the proportion of bird visitation events (based on expert manual annotation; Method 1) that was captured (true positive) or missed (false negative) by the automatic method. All the proportions were calculated for each video, then subsequently averaged across videos.

Results

Automatic method is comparable to manual annotation

We found a significant positive correlation between expert manual annotation and the automatic method ($r = 0.62$, 95% CI: 0.58–0.66, $p < 0.001$; Fig. 2a). We also found that both proxies of parental care behaviour significantly predicted an increase in both the number of fledglings and recruits in our statistical analysis (Fig. 3). None of the fixed and random effects predicted fitness outcomes, except for a significant negative quadratic effect for hatch date in most models, showing that fitness is maximised in the mid-breeding season (Supporting information). To test for the effect of outliers, we also ran the same models after removing data from nests with five recruits or fledglings, which yielded similar results (Supporting information).

Comparing data collection methods

The proportion for the automatic method was significantly inflated based on the proportion we computed for the number of detected events compared to the baseline (Table 2, Fig. 2b). The proportion for crowd-sourcing and hybrid annotation was not significantly different from the baseline (Table 2, Fig. 2b), showing that the data obtained were consistent with expert annotation. The estimates obtained by the crowd-sourcing undergraduate students were consistent, with the majority of students measuring a feeding rate within a difference of 10 compared to the baseline value (relative to mean of ~ 58.5 events in each video; Supporting information). On average, the crowd-sourcing data extraction by undergraduate students took 65.4 min per video (min 25 min, max 100 min), whereas the hybrid expert annotation method took an average of 12.0 min per video (min 4.6 min, max 31.0 min). This was an average of 53 min less time required when first processing videos using the hybrid pipeline, but more time consuming overall (102; $12 + \sim 90$ min computer processing time).

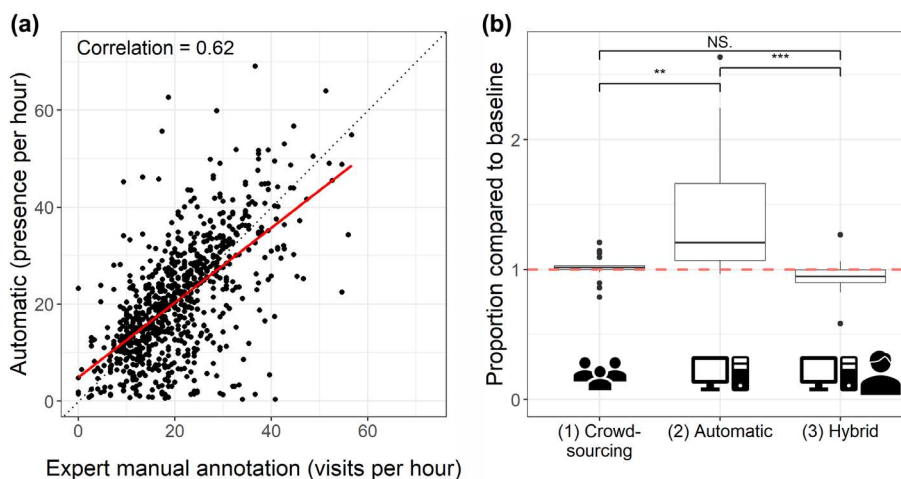


Figure 2. Comparison of data collection methods for parental provisioning videos of house sparrows *Passer domesticus* on Lundy Island, UK. (a) Positive correlation between the presence and visit rates collected by the automatic method and expert manual annotation. The red line shows a significant positive correlation, and the dotted line shows a 1:1 line if both rates were equal. (b) Comparison of the proportion of events collected compared to the expert annotation baseline of the 18-video subset. (1) Crowd-sourcing : annotation by a cohort of undergraduate students. (2) Automatic: collected by automatic pipeline without human intervention. (3) Hybrid: expert manual annotation of output clips identified by the automatic pipeline. The red dotted line represents the reference level of 1, brackets and labels represent the significance of paired *t*-tests between each pair of datasets (NS: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

To gain further insight into the sources of error and performance of the automatic method, we computed a confusion matrix using data from hybrid manual annotation (Table 3). We show that the automatic pipeline had a false negative rate of 0.104. However, there was a higher false positive rate of 0.260, when many non-visitation events were captured. Further analysis showed that only 0.044 of visits were true false detections where no birds were present, and the majority (0.216) were caused by non-visitation events, where the birds were detected correctly by DeepMeerkat, but the bird was not visiting the nest.

Discussion

In this study, we set out to compare four different methods for collecting parental provisioning data on a pre-existing video dataset of house sparrows to determine whether available open-source software can aid biologists in laborious data collection tasks. Using an automatic data collection method based on DeepMeerkat (Weinstein 2018b), we extracted visitation rates from parental provisioning videos of house sparrows and found that the results obtained correlated positively with expert manual annotation. Our data analysis also shows

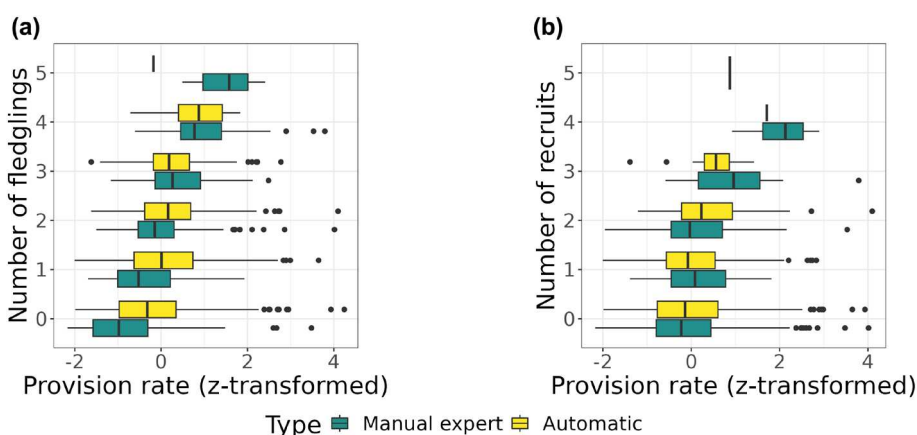


Figure 3. Relationship between the number of (a) fledglings and (b) recruits with mean provisioning rates in house sparrows *Passer domesticus* on Lundy Island, UK. Provisioning rates (visits/h) for the manual expert method (blue) were collected by trained individuals, while provisioning rates (presence/h) for automatic methods (yellow) were obtained by the machine learning method without human intervention. Provisioning rates were further z-transformed to allow effect sizes to be comparable, with 0 representing the mean and each unit representing one standard deviation away from the mean.

Table 2. Comparison of three annotation methods using a validation dataset of 18 provisioning videos of house sparrows *Passer domesticus* on Lundy Island, UK. Proportions compared to the baseline were calculated by dividing the number of detected events using each method by the measure obtained by expert manual annotation. Test statistics were obtained from one-sample *t*-tests, with the theoretical mean (μ) set to 1.

| Method | Mean | 95% Confidence interval | | <i>t</i> | p-value |
|-------------------------------|------|-------------------------|----------|----------|---------|
| | | Mean | interval | | |
| (1) Crowd-sourcing annotation | 1.01 | 0.97–1.06 | | 1.13 | 0.52 |
| (2) Automatic | 1.39 | 1.17–1.63 | | 3.68 | 0.002 |
| (3) Hybrid annotation | 0.94 | 0.88–1.01 | | −1.78 | 0.10 |

that the automatic method led to comparable predictions of increased fitness for parents that expressed more parental care behaviour, showing that this new metric can be used as a proxy of parental investment (Trivers 1972, Schroeder et al. 2013). However, further analysis shows that the automatic method has high false positive and false negative error rates, since DeepMeerkat was designed to detect the presence of animals (Weinstein 2018b), and not the specific behaviour of feeding the young. We highlight that a hybrid approach in which raw videos are first processed through DeepMeerkat to obtain clips, then manually annotated, can obtain estimates equivalent to expert manual annotation in approximately 20% of the time. With the rapid development of open-source tools, we encourage any researchers with large video datasets to evaluate currently available software and, if required, to apply a hybrid method for annotation to considerably reduce the time investment needed to process huge backlogs of video data if there is no trade-off in accuracy.

Using the open-source software DeepMeerkat, we estimated an automatic presence rate that broadly correlated with expert manual annotation ($r=0.62$) and predicts an increase in fledglings and recruits in broods. However, while a fully automatic method seems to be able to extract biologically meaningful presence rates for the Lundy sparrows, the effect size is quite low and should be interpreted with

Table 3. Confusion matrix to evaluate performance of the automatic method (labelled 'Predicted value') compared with hybrid annotation by experts (labelled 'Actual value'), based on 18 provisioning videos of house sparrows *Passer domesticus* on Lundy Island, UK. Reported as a mean (standard deviation). ¹Proportion of nest visitation events detected by automatic method. ²Proportion of false detections by automatic method. These are further separated into: No presence: no bird was present in frame; No visitation: bird in frame, but not visiting nest. ³Proportion of bird visitation events missed entirely by automatic method. ⁴No estimate for true negative rates.

| | | Actual value | |
|------------------------|----------|--|--|
| | | Positive | Negative |
| Predicted value | Positive | True positive ¹ 0.896 (0.11) | False positive ² 0.260 (0.16) No presence: 0.044 (0.07) No visitation: 0.216 (0.14) |
| | Negative | False negative ³ 0.104 (0.11) | True negative ⁴ NA |

caution. This result may also not generalize well, because the current study system had a large sample size and a broad correlation between feeding bouts and the frequency of parents being in proximity with the nest (as a weaker alternative measure of parental care). The placement of the camera to collect these videos also contributed to the effectiveness of DeepMeerkat, since it was pointed directly at the nest box, and most movements in the frame were caused by parents visiting the nest rather than being movements in the environment or by chicks. As such, DeepMeerkat was an appropriate tool to use given the current dataset, but may be less applicable to open nest systems (Reif and Tornberg 2006) or cameras placed inside nest boxes (Zárybnická et al. 2016), which can dramatically increase the number of false detections from the movement of chicks or the background. This was also evident from the small subset of videos (1.31%) that was removed in the current dataset due to over-detection. To have an automatic pipeline that can reduce these biases, alternative custom machine learning algorithms will need to be designed to automatically recognize nest visits instead of just bird presence, using behavioural classification techniques (Conway et al. 2021, Ditria et al. 2021). However, this was largely outside the scope of the current study and represents an exciting future development.

While we show that DeepMeerkat was appropriate for the current study system, further analysis of the data collected by the hybrid annotation method revealed various sources of error when applying the automatic pipeline. We showed that the pipeline produced biased results, with a false negative rate averaging 10%, showing that the software can miss visitation events. However, this under-detection was negated by a high false positive rate (26%), mainly caused by the over-detection and misclassification of non-visitation events (22%) and false detections (4%), causing the overall estimate to be over-inflated. The larger false positive rate can be attributed to the original design of DeepMeerkat, which is intended to detect movement between frames, instead of feeding behaviours. In this sense, fine-tuning DeepMeerkat using more frames from the current study system may potentially have reduced the false negative and false detection rates, but would have been ineffective because the inflated measure was mainly due to misclassification of non-visitation events (22%). Another source of the inflated measure can also be attributed to the definition of 7 s for each event, where events can be over-inflated when parents are staying in frame for a longer period of time. Better clustering algorithms can be applied to group events without relying on simple thresholds. Overall, DeepMeerkat was an appropriate choice of software for the current system, even though it produces biased estimates. While alternative approaches such as training an object detection model (Jocher et al. 2023) might reduce false detections, it will still result in similarly inflated measures unless a model is explicitly trained to detect behaviours.

Using a smaller subset of testing videos, we show that crowd-sourcing annotation and hybrid expert annotation were both comparable with the baseline set by expert manual annotation. Of all the methods, the crowd-sourcing

annotation appeared to be the most accurate, albeit with the recorded proportion of events compared to the baseline being larger than one, suggesting that the undergraduate students may have identified some visitation events that were missed by expert manual annotation. This is possibly due to the multi-observer effect (Guay et al. 2013). We also show that the variation in parental visitation estimates for each individual annotator were consistent, within ± 10 visitation events compared to the baseline, with a mean of ~ 58.5 events per video (Supporting information). However, the between-annotator variance of 10 visits can still introduce bias in provisioning estimates, hence we still recommend multiple observers to annotate a single video for future researchers who look to adopt a crowd-sourcing approach.

Even though crowd-sourcing seems to be the most accurate method, the annotation time for hybrid annotation by the investigator only takes approximately 20% of the time compared to crowd-sourcing annotation, and may still obtain comparable and accurate visitation rates. While it takes more time overall once computing processing is considered (average ~ 102 min compared to 65.4 min), computational time is much cheaper than human labour, and videos can be processed non-stop (e.g. overnight). For example, if there are 100 videos each 90 min in length in a single field season, all the data can be realistically processed on one laptop in approximately 9000 min, which is approximately 6.25 days, i.e. just under a week non-stop. We also show that the under-detection (false negative rate: 10%) reported above is negligible, such that after manual annotation in the hybrid method, the estimates of provisioning rates are still comparable to the baseline. A weakness of the approach is that if researchers would like to retrieve these missed detections, they will have to review the whole video from scratch, so future researchers will have to be aware that the parental visit rates estimated by this method might be potentially biased. Researchers who adopt this workflow in their study systems should also first make an evaluation and ensure the false negative rate is negligible, similar to the current case study. Nevertheless, we highlight the value of first using an open-source machine learning tool to pre-process videos for annotations, which not only reduces the actual annotation time, but also the time needed to train observers in the appropriate experimental protocol.

For field protocols that involve video recordings, we encourage researchers to explore available open-source machine learning tools and to evaluate the methods against traditional protocols. If the automatic tools produce comparable estimates, the method can be used directly to automate data collection, but if the pipeline produces biased estimates, researchers can consider hybrid approaches to save annotation time. For researchers with similarly large time-lapse video datasets in the field, we also recommend the use of DeepMeerkat, instead of custom algorithms that require more domain-specific knowledge. DeepMeerkat not only has a user-friendly graphical interface but also requires limited investment in terms of machine learning knowledge or expensive hardware (e.g. GPUs).

Conclusion

Machine learning and computer vision approaches are becoming widely used in ecology (Borowiec et al. 2021, Couzin and Heins 2022), but adoption of these frameworks by ornithologists is still slow and limited. Here we present a case study of using an open-source software to first pre-process long-duration videos before annotation, to considerably reduce annotation time without significant reduction in accuracy. With the increase in available open-source tools to reduce manual annotation efforts (Van Horn et al. 2015, Weinstein 2018b, Lauer et al. 2021, Walter and Couzin 2021), and the increase in computing literacy of ecology graduates (Farrell and Carey 2018), we encourage researchers to evaluate the applicability of these tools and, if needed, make use of hybrid approaches. This would not only unlock the major time bottleneck of unanalysed data that would otherwise go to waste, but also allow more interesting biological hypotheses to be tested.

Acknowledgements – We would like to thank the Lundy Landmark Trust and the Lundy Field Society for their ongoing support for our fieldwork.

Funding – This research was supported by Imperial College London. JS was awarded a fellowship from the Volkswagen Foundation; a grant from the German Research Foundation: Deutsche Forschungsgemeinschaft; and grant no. CIG PCIG12-GA-2012-333096 from the European Research Council. TB was awarded grant no. NE/J024597/1. WDP and the Pearse lab were funded by NSF grant no. ABI-1759965 and by the UKRI/NERC NE/V009710/1. AHHC was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2117 – grant no. 422037984.

Permits – Fieldwork was carried out, including tissue collection for DNA extraction, with the permission of the Lundy Company and Field Society and under permits from the UK Home Office (PP7009092 and PP5873078) and BTO (S:6308).

Author contributions

Alex Hoi Hang Chan: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review and editing (equal). **Jingqi Liu:** Data curation (equal); Methodology (equal); Visualization (equal); Writing – review and editing (equal). **Terry Burke:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Writing – review and editing (equal). **William D. Pearse:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – review and editing (equal). **Julia Schroeder:** Conceptualization (equal); Data curation (equal); Funding acquisition (equal); Investigation (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/jav.03167>.

Data availability statement

Code and data are available from the Zenodo Repository: <https://doi.org/10.5281/zenodo.6411359> (Chan et al. 2023). The automatic processing pipeline is available from the Zenodo Digital Repository: <https://zenodo.org/records/10190356>. We provide a sample script in R to post-process outputs of DeepMeerkat into short video clips.

Supporting information

The Supporting information associated with this article is available with the online version.

References

- Bogucki, R., Cygan, M., Khan, C. B., Klimek, M., Milczek, J. K. and Mucha, M. 2019. Applying deep learning to right whale photo identification. – *Conserv. Biol.* 33: 676–684.
- Borowiec, M. L., Frandsen, P. B., Dikow, R. B., McKeeken, A., Valentini, G. and White, A. E. 2021. Deep learning as a tool for ecology and evolution. – *Methods Ecol. Evol.* 13: 1640–1660.
- Bradski, G. 2000. The openCV library. – *Dr Dobbs J. Softw. Tools* 120: 122–125.
- Chan, A. H. H., Liu, J., Burke, T., Pearse, W. D. and Schroeder, J. 2023. Data from: Comparison of manual, machine learning, and hybrid methods for video annotation to extract parental care data. – Zenodo Digital Repository, <https://doi.org/10.5281/zenodo.6411359>.
- Cleasby, I. R., Burke, T., Schroeder, J. and Nakagawa, S. 2011. Food supplements increase adult tarsus length, but not growth rate, in an island population of house sparrows (*Passer domesticus*). – *BMC Res. Notes* 4: 431, <https://doi.org/10.1186/1756-0500-4-431>.
- Clutton-Brock, T. H. 1991. The evolution of parental care. – Princeton Univ. Press.
- Cockburn, A. 2006. Prevalence of different modes of parental care in birds. – *Proc. R. Soc. B.* 273: 1375–1383.
- Conway, A. M., Durbach, I. N., McInnes, A. and Harris, R. N. 2021. Frame-by-frame annotation of video recordings using deep neural networks. – *Ecosphere* 12: e03384.
- Couzin, I. D. and Heins, C. 2022. Emerging technologies for behavioral research in changing environments. – *Trends Ecol. Evol.* 38: 346–354.
- Cresswell, W. and McCleery, R. 2003. How great tits maintain synchronization of their hatch date with food supply in response to long-term variability in temperature. – *J. Anim. Ecol.* 72: 356–366.
- Desell, T., Goehner, K., Andes, A., Eckroad, R. and Ellis-Felege, S. 2015. On the effectiveness of crowd sourcing avian nesting video analysis at Wildlife@Home. – *Proc. Comput. Sci.* 51: 384–393.
- Ditria, E. M., Jinks, E. L. and Connolly, R. M. 2021. Automating the analysis of fish grazing behaviour from videos using image classification and optical flow. – *Anim. Behav.* 177: 31–37.
- Dunn, P. O. and Cockburn, A. 1996. Evolution of male parental care in a bird with almost complete cuckoldry. – *Evolution* 50: 2542–2548.
- Estramil, N., Eens, M. and Müller, W. 2013. Coadaptation of offspring begging and parental provisioning - an evolutionary ecological perspective on avian family life. – *PLoS One* 8: e70463.
- Farrell, K. J. and Carey, C. C. 2018. Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. – *Ecol. Evol.* 8: 7744–7751.
- Ferreira, A. C., Silva, L. R., Renna, F., Brandl, H. B., Renoult, J. P., Farine, D. R., Covas, R. and Doutrelant, C. 2020. Deep learning-based methods for individual recognition in small birds. – *Methods Ecol. Evol.* 11: 1072–1085.
- García-Navas, V. and Sanz, J. J. 2010. Flexibility in the foraging behavior of blue tits in response to short-term manipulations of brood size. – *Ethology* 116: 744–754.
- Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R. and Couzin, I. D. 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. – *eLife* 8: e47994.
- Guay, P.-J., McLeod, E. M., Cross, R., Formby, A. J., Maldonado, S. P., Stafford-Bell, R. E., St-James-Turner, Z. N., Robinson, R. W., Mulder, R. A. and Weston, M. A. 2013. Observer effects occur when estimating alert but not flight-initiation distances. – *Wildl. Res.* 40: 289–293.
- Hadfield, J. 2014. MCMCglmm course notes. – <https://cran.r-project.org/web/packages/MCMCglmm/MCMCglmm.pdf>.
- Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. – *J. Stat. Softw.* 33: 1–22.
- Harmer, A. M. T. and Thomas, D. B. 2019. pathtrackr: an R package for video tracking and analysing animal movement. – *Methods Ecol. Evol.* 10: 1196–1202.
- Ihle, M., Pick, J. L., Winney, I. S., Nakagawa, S., Schroeder, J. and Burke, T. 2019. Rearing success does not improve with apparent pair coordination in offspring provisioning. – *Front. Ecol. Evol.* 7: 405.
- Jocher, G., Chaurasia, A. and Qiu, J. 2023. YOLO by ultralytics (Version 8.0.0). – <https://github.com/ultralytics/ultralytics>.
- Körschens, M., Barz, B. and Denzler, J. 2018. Towards automatic identification of elephants in the wild. – *ArXiv181204418*.
- Lack, D. L. 1968. Ecological adaptations for breeding in birds. – Methuen.
- Lattore, M., Nakagawa, S., Burke, T., Plaza, M. and Schroeder, J. 2019. No evidence for kin recognition in a passerine bird. – *PLoS One* 14: e0213486.
- Lauer, J., Zhou, M., Ye, S., Menegas, W., Schneider, S., Nath, T., Rahman, M. M., Di Santo, V., Soberanes, D., Feng, G., Murthy, V. N., Lauder, G., Dulac, C., Mathis, M. W. and Mathis, A. 2022. Multi-animal pose estimation, identification and tracking with DeepLabCut. – *Nat. Methods* 19: 496–504.
- Lendvai, Á. Z., Akçay, Ç., Ouyang, J. Q., Dakin, R., Domalik, A. D., St John, P. S., Stanback, M., Moore, I. T. and Bonier, F. 2015. Analysis of the optimal duration of behavioral observations based on an automated continuous monitoring system in tree swallows (*Tachycineta bicolor*): is one hour good enough? – *PLoS One* 10: e0141194.
- Magnussen, E. and Jensen, J.-K. 2017. Ringing recoveries of house sparrow (*Passer domesticus*) in the Faroe Islands during the years 1963–2007. – *Fróðskaparrit-Faroese Sci. J.* 57: 182–189.
- Marcot, B. G., Lorenz, T. J., Fischer, P., Weinstein, B. G. and Cowell, S. 2019. Efficacy of automated detection of motion in wildlife monitoring videos. – *Wildl. Soc. Bull.* 43: 726–736.

- Mariette, M. M., Pariser, E. C., Gilby, A. J., Magrath, M. J. L., Pryke, S. R. and Griffith, S. C. 2011. Using an electronic monitoring system to link offspring provisioning and foraging behavior of a wild passerine. – *Auk* 128: 26–35.
- Mertens, J. E., Brisson, L., Janeček, Š., Klomberg, Y., Maicher, V., Sáfán, S., Delabye, S., Potocký, P., Kobe, I. N. and Pycrz, T. 2021. Elevational and seasonal patterns of butterflies and hawkmoths in plant-pollinator networks in tropical rainforests of Mount Cameroon. – *Sci. Rep.* 11: 1–12.
- Mitchell, G. W., Newman, A. E., Wikelski, M. and Ryan Norris, D. 2012. Timing of breeding carries over to influence migratory departure in a songbird: an automated radiotracking study. – *J. Anim. Ecol.* 81: 1024–1033.
- Nakagawa, S., Gillespie, D. O. S., Hatchwell, B. J. and Burke, T. 2007. Predictable males and unpredictable females: sex difference in repeatability of parental care in a wild bird population. – *J. Evol. Biol.* 20: 1674–1681.
- Owens, I. P. and Bennett, P. M. 1994. Mortality costs of parental care and sexual dimorphism in birds. – *Proc. R. Soc. B* 257: 1–8.
- Pearse, W. D., Morales-Castilla, I., James, L. S., Farrell, M., Boivin, F. and Davies, T. J. 2018. Global macroevolution and macroecology of passerine song. – *Evolution* 72: 944–960.
- Pegoraro, L., Hidalgo, O., Leitch, I. J., Pellicer, J. and Barlow, S. E. 2020. Automated video monitoring of insect pollinators in the field. – *Emerg. Top. Life Sci.* 4: 87–97.
- Pennington, Z. T., Dong, Z., Feng, Y., Vetere, L. M., Page-Harley, L., Shuman, T. and Cai, D. J. 2019. ezTrack: an open-source video analysis pipeline for the investigation of animal behavior. – *Sci. Rep.* 9: 19979.
- Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M. and Shaevitz, J. W. 2019. Fast animal pose estimation using deep neural networks. – *Nat. Methods* 16: 117–125.
- Priyadarshani, N., Marsland, S. and Castro, I. 2018. Automated birdsong recognition in complex acoustic environments: a review. – *J. Avian Biol.* 49: jav-01447.
- Reif, V. and Tornberg, R. 2006. Using time-lapse digital video recording for a nesting study of birds of prey. – *Eur. J. Wildl. Res.* 52: 251–258.
- Ringsby, T. H., Berge, T., Saether, B.-E. and Jensen, H. 2009. Reproductive success and individual variation in feeding frequency of house sparrows (*Passer domesticus*). – *J. Ornithol.* 150: 469–481.
- Robie, A. A., Seagraves, K. M., Egnor, S. E. and Branson, K. 2017. Machine vision methods for analyzing social interactions. – *J. Exp. Biol.* 220: 25–34.
- Root-Gutteridge, H., Brown, L. P., Forman, J., Korzeniowska, A. T., Simner, J. and Reby, D. 2021. Using a new video rating tool to crowd-source analysis of behavioural reaction to stimuli. – *Anim. Cogn.* 24: 947–956.
- Rose, A. P. 2009. Temporal and individual variation in offspring provisioning by tree swallows: a new method of automated nest attendance monitoring. – *PLoS One* 4: e4111.
- Rose, K. E., Clutton-Brock, T. H. and Guinness, F. E. 1998. Cohort variation in male survival and lifetime breeding success in red deer. – *J. Anim. Ecol.* 67: 979–986.
- Royle, N. J., Smiseth, P. T. and Kölliker, M. 2012. The evolution of parental care. – Oxford Univ. Press.
- Sánchez-Tójar, A., Winney, I., Girndt, A., Simons, M. J. P., Nakagawa, S., Burke, T. and Schroeder, J. 2017. Winter territory prospecting is associated with life-history stage but not activity in a passerine. – *J. Avian Biol.* 48: 407–416.
- Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D. and Carvalho, S. 2019. Chimpanzee face recognition from videos in the wild using deep learning. – *Sci. Adv.* 5: eaaw0736.
- Schroeder, J., Cleasby, I. R., Nakagawa, S., Ockendon, N. and Burke, T. 2011. No evidence for adverse effects on fitness of fitting passive integrated transponders (PITs) in wild house sparrows *Passer domesticus*. – *J. Avian Biol.* 42: 271–275.
- Schroeder, J., Nakagawa, S., Cleasby, I. R. and Burke, T. 2012. Passerine birds breeding under chronic noise experience reduced fitness. – *PLoS One* 7: e39200.
- Schroeder, J., Cleasby, I., Dugdale, H. L., Nakagawa, S. and Burke, T. 2013. Social and genetic benefits of parental investment suggest sex differences in selection pressures. – *J. Avian Biol.* 44: 133–140.
- Schroeder, J., Nakagawa, S., Rees, M., Mannarelli, M. E. and Burke, T. 2015. Reduced fitness in progeny from old parents in a natural population. – *Proc. Natl Acad. Sci. USA* 112: 4021–4025.
- Schroeder, J., Hsu, Y. H., Winney, I., Simons, M., Nakagawa, S. and Burke, T. 2016. Predictably philandering females prompt poor paternal provisioning. – *Am. Nat.* 188: 219–230.
- Schroeder, J., Dugdale, H., Nakagawa, S., Sparks, A. and Burke, T. 2019. Social genetic effects (IGE) and genetic intra- and inter-sexual genetic correlation contribute to the total heritable variance in parental care. – *EcoEvoRxiv*, <https://ecoevorxiv.org/repository/view/4513/>.
- Sheehan, E. V., Bridger, D., Nancollas, S. J. and Pittman, S. J. 2020. PelagiCam: a novel underwater imaging system with computer vision for semi-automated monitoring of mobile marine fauna at offshore structures. – *Environ. Monit. Assess.* 192: 1–13.
- Simons, M. J. P., Winney, I., Nakagawa, S., Burke, T. and Schroeder, J. 2015. Limited catching bias in a wild population of birds with near-complete census information. – *Ecol. Evol.* 5: 3500–3506, <https://doi.org/10.1002/ece3.1623>.
- Sridhar, V. H., Roche, D. G. and Giggins, S. 2019. Tractor: image-based automated tracking of animal movement and behaviour. – *Methods Ecol. Evol.* 10: 815–820.
- Stearns, S. 1992. The evolution of life histories. – Oxford Univ. Press.
- Trivers, R. 1972. Parental investment and sexual selection. – In: Campbell, B. (ed.), *Sexual selection and the descent of man*. Aldine, pp. 136–179.
- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F. and Burghardt, T. 2022. Perspectives in machine learning for wildlife conservation. – *Nat. Commun.* 13: 1–15.
- Tuytens, F. A. M., de Graaf, S., Heerkens, J. L. T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E. and Ampe, B. 2014. Observer bias in animal behaviour research: can we believe what we score, if we score what we believe? – *Anim. Behav.* 90: 273–280.
- Unger, S. 2022. Power of a snapshot observation: using iNaturalist to teach undergraduates about ethograms in animal behavior. – *J. Microbiol. Biol. Educ.* 24: e00044-22.
- Valletta, J. J., Torney, C., Kings, M., Thornton, A. and Madden, J. 2017. Applications of machine learning in animal behaviour studies. – *Anim. Behav.* 124: 203–220.
- Van Horn, G., Branson, S., Farrell, R., Haber, S., Barry, J., Ipeiro-tis, P., Perona, P. and Belongie, S. 2015. Building a bird recognition app and large scale dataset with citizen scientists: the fine

- print in fine-grained dataset collection. – In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 595–604.
- Voss, M. A. and Cooper, C. B. 2010. Using a free online citizen-science project to teach observation & quantification of animal behavior. – *Am. Biol. Teach.* 72: 437–443.
- Wäldchen, J. and Mäder, P. 2018. Machine learning for image based species identification. – *Methods Ecol. Evol.* 9: 2216–2225.
- Walter, T. and Couzin, I. D. 2021. TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. – *eLife* 10: e64000.
- Weinstein, B. G. 2015. MotionMeerkat: integrating motion video detection and ecological monitoring. – *Methods Ecol. Evol.* 6: 357–362.
- Weinstein, B. G. 2018a. A computer vision for animal ecology. – *J. Anim. Ecol.* 87: 533–545.
- Weinstein, B. G. 2018b. Scene-specific convolutional neural networks for video-based biodiversity detection. – *Methods Ecol. Evol.* 9: 1435–1441.
- Westneat, D. F., Bókony, V., Burke, T., Chastel, O., Jensen, H., Kvalnes, T., Lendvai, Á. Z., Liker, A., Mock, D., Schroeder, J., Schwagmeyer, P. L., Sorci, G. and Stewart, I. R. K. 2014. Multiple aspects of plasticity in clutch size vary among populations of a globally distributed songbird. – *J. Anim. Ecol.* 83: 876–887.
- Wiebe, K. L. 2018. Age-related improvements in fecundity are driven by the male in a bird with partially reversed sex roles in parental care. – *Oecologia* 188: 1095–1104.
- Wilcoxon, T. E., Boughton, R. K. and Schoech, S. J. 2010. Older can be better: physiological costs of paternal investment in the Florida scrub-jay. – *Behav. Ecol. Sociobiol.* 64: 1527–1535.
- Winkler, D. W. and Wilkinson, G. S. 1988. Parental effort in birds and mammals: theory and measurement. – In: Harvey, P. H. and Partridge, L. (eds), *Oxford surveys in evolutionary biology*. Oxford Univ. Press, pp. 185–214.
- Winney, I., Nakagawa, S., Hsu, Y.-H., Burke, T. and Schroeder, J. 2015. Troubleshooting the potential pitfalls of cross-fostering. – *Methods Ecol. Evol.* 6: 584–592.
- Wojczulanis-Jakubas, K., Araya-Salas, M. and Jakubas, D. 2018. Seabird parents provision their chick in a coordinated manner. – *PLoS One* 13: e0189969.
- Zárybnická, M., Kubizňák, P., Šindelář, J. and Hlaváč, V. 2016. Smart nest box: a tool and methodology for monitoring of cavity-dwelling animals. – *Methods Ecol. Evol.* 7: 483–492.