RESEARCH ARTICLE

Contemporary Methods for Studying Animal Sociality in the Wild

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Automated face recognition using deep neural networks produces robust primate social networks and sociality measures

Daniel P. Schofield[1,2] | Gregory F. Albery[3,4] | Josh A. Firth[5] | Alexander Mielke[6] | Misato Hayashi[7,8] | Tetsuro Matsuzawa[7,9,10] | Dora Biro[5,11] | Susana Carvalho[1,12,13]

[1]Primate Models for Behavioural Evolution Lab, Institute of Human Sciences, University of Oxford, Oxford, UK; [2]Visual Geometry Group, Department of Engineering Science, University of Oxford, Oxford, UK; [3]Department of Biology, Georgetown University, Washington, DC, USA; [4]Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany; [5]Department of Biology, University of Oxford, Oxford, UK; [6]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK; [7]Chubu Gakuin University, Seki, Japan; [8]Japan Monkey Centre, Inuyama, Japan; [9]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California, USA; [10]College of Life Sciences, Northwest University, Xi'an, China; [11]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York, USA; [12]Interdisciplinary Center for Archaeology and Evolution of Human Behaviour (ICArEHB), Universidade do Algarve, Faro, Portugal and [13]Gorongosa National Park, Chitengo, Mozambique

Correspondence
Daniel P. Schofield
Email: daniel.schofield@eng.ox.ac.uk

Abstract

1. Longitudinal video archives of behaviour are crucial for examining how sociality shifts over the lifespan in wild animals. New approaches adopting computer vision technology hold serious potential to capture interactions and associations between individuals in video at large scale; however, such approaches need a priori validation, as methods of sampling and defining edges for social networks can substantially impact results.

2. Here, we apply a deep learning face recognition model to generate association networks of wild chimpanzees using 17 years of a video archive from Bossou, Guinea. Using 7 million detections from 100 h of video footage, we examined how varying the size of fixed temporal windows (i.e. aggregation rates) for defining edges impact individual-level gregariousness scores.

3. The highest and lowest aggregation rates produced divergent values, indicating that different rates of aggregation capture different association patterns. To avoid any potential bias from false positives and negatives from automated detection, an intermediate aggregation rate should be used to reduce error across multiple variables. Individual-level network-derived traits were highly repeatable, indicating strong inter-individual variation in association patterns across years and highlighting the reliability of the method to capture consistent individual-level patterns of sociality over time. We found no reliable effects of age and sex on social behaviour and despite a significant drop in population size over the study period, individual estimates of gregariousness remained stable over time.

4. We believe that our automated framework will be of broad utility to ethology and conservation, enabling the investigation of animal social behaviour from video footage at large scale, low cost and high reproducibility. We explore the implications of our findings for understanding variation in sociality patterns in wild ape populations. Furthermore, we examine the trade-offs involved in using face recognition technology to generate social networks and sociality measures. Finally, we outline the steps for the broader deployment of this technology for analysis of large-scale datasets in ecology and evolution.

1 | INTRODUCTION

Social behaviour has important implications for survival and reproduction, and understanding sociality is therefore important for understanding animals' evolution and ecology (Abbot & Rubenstein, 2017; Krause & Ruxton, 2002). Using interaction patterns and spatial associations between group members, biologists can gain insight into the structure, function and evolutionary consequences of sociality for animals in the wild (Alexander, 1974; Hinde, 1976; Silk & Kappeler, 2017). Social network analysis has proven to be a useful tool for quantifying individual-level patterns of association and group-level social structure in a wide array of species (Croft et al., 2008; Krause et al., 2015), including birds (Aplin, Farine, et al., 2015; St Clair et al., 2015), cetaceans (Allen et al., 2013) and primates (Sueur et al., 2011). Long-term field sites and repeated observations of individuals across many years are required to capture the demographic fluctuation of populations and the ecological and evolutionary processes driving species' adaptations and social behaviour (Clutton-Brock & Sheldon, 2010). Large-scale image datasets, in particular remote camera trap databases and video archives (Burton et al., 2015; O'Connell et al., 2010) at high temporal resolution, have potential to examine behavioural change over time. Conventional methods employ human researchers to code behavioural information; however, this is becoming increasingly difficult to scale as video databases accumulate, limiting the scope of research and the statistical power of analysis. For typical coding tasks such as individual identification from camera trap footage, learning to reliably identify habituated individuals can take a matter of weeks for experienced coders, whilst training inexperienced coders or for more challenging tasks such as labelling unhabituated individuals or multiple behavioural variables, it can take months of research effort (McCarthy et al., 2019).

Recent developments in artificial intelligence (AI) and the field of computer vision, boosted by advances in deep learning (LeCun et al., 2015), have opened up the possibility for automating behavioural analysis from video (Vidal et al., 2021; Weinstein, 2018). Individual detection, tracking and recognition using deep learning can be used to measure patterns of association and social networks

at high frame rates (Schofield et al., 2019). Automated micro-sensing technologies such as animal-borne radio frequency identification tags (Firth & Sheldon, 2016; Gelardi et al., 2020), proximity loggers (St Clair et al., 2015) and GPS trackers (Strandburg-Peshkin et al., 2015) can track animal behaviour at high resolution and accuracy, but require individual animals to be fitted with devices, which introduces logistical and ethical challenges that can be prohibitively invasive for some species such as chimpanzees (Dore et al., 2020). The use of deep learning for individual identification (Chen et al., 2020; Ferreira, Silva, et al., 2020) can efficiently generate sociality information for a wider array of species that are difficult to observe in situ or cannot be tagged or tracked using biologgers. Visual data from archives are valuable as they can reveal the nature and behavioural context of individual interactions. Video data can be collected through different methods such as systematic remotely operated camera trap grids or hand operated video cameras. Although the use of video offers a valuable way to capture a detailed record of events, it also presents several visual challenges. For example, remote camera traps, which are fixed in situ, can miss animal interactions due to their limited fields of vision. Hand-operated cameras can move more freely to capture social behaviours by panning or zooming, but this movement can cause motion blur or introduce observer bias by focusing on specific individuals or behaviours. Additionally, video data collected in the wild often has low resolution, poor lighting, and may be occluded by vegetation, making it a challenge to capture animal associations accurately and without bias.

Once video has been collected, behavioural information, such as the distribution of individuals in video across time and space, can be automatically extracted using individual recognition classifiers to produce a time series of individual detections and co-presence (Schofield et al., 2019). The data can be summarised to capture social relationships by generating 'association'-based social networks, recording 'co-occurrences' if individuals appear together at a set location, within a certain threshold of distance, or within a time window, as an indication of the rate of interaction and the strength of the relationship—these relationships are represented as 'edges'/edge weights between 'nodes' in a network (Farine & Whitehead, 2015). Automating data collection of social behaviour through video-based

analysis using AI enables high-throughput sampling compared to coding this information manually from in situ observation or retrospectively from video records, and thus provides greater statistical power which is required for robust estimates of these social indices (Franks et al., 2010; James et al., 2009). However, social interaction is highly dynamic, and choosing a set threshold for defining 'grouping events' of edges/edge weights between individuals in a network (otherwise referred to here as the aggregation rate, that is the size of the temporal window used for defining edges/co-occurrences and aggregating individual associations) to detect the 'real' biological signal for capturing meaningful associations is a considerable challenge (Ferreira, Covas, et al., 2020); furthermore, these thresholds are likely to be species, population specific and may need to be adjusted depending on the study design or context. Thus, far few studies have examined how the choice of sampling strategies, methods of data collection and network aggregation can impact biological results (Castles et al., 2014; Davis et al., 2018; Farine, 2015; Gelardi et al., 2020; He et al., 2022). Choosing a fixed threshold for the size of temporal window used introduces trade-offs which need further investigation. For example, using high thresholds/short time windows (e.g. recording interactions every second) increases data throughput, but can aggregate too much noise (Farine, 2018) and may introduce false negatives by failing to detect social interactions (Type II error). Conversely, using low thresholds/aggregation rates with larger time windows may increase false positives by aggregating individuals who were not interacting (Type I error), thereby producing networks which are lacking in variation and too low in resolution (Wang et al., 2014). The aim is to record associations at a high enough temporal resolution to capture meaningful variation in sociality without oversampling and introducing errors to the social networks.

In this study, we use a long-term video archive of wild chimpanzees *Pan troglodytes verus* from Bossou in Guinea to investigate the reliability of face recognition for producing social networks and sociality estimates at the individual level. To assess potential biases from use of human-operated cameras in our dataset (e.g. individuals missed out of shot), we examine how varying the size of fixed temporal windows for sampling grouping events affects the patterns and consistency of social behaviour estimates over time. Bossou offers a unique opportunity to examine social behaviour in the context of tool use, where chimpanzees have been continuously observed over 30 years at a clearing in the forest to crack nuts using stone tools (Matsuzawa, 1994; Sugiyama & Koman, 1979). Using footage recorded annually at the so-called 'outdoor laboratory', we examine three measures of gregariousness in chimpanzees: (a) individual social network centrality (Strength); (b) the average size of the party in which an individual is observed (Mean Group Size) and (c) the proportion of time spent alone (Time Alone). The main aims of this paper are to (1) investigate how the aggregation rate/size of the temporal window used for capturing co-occurrences affects these estimates; (2) investigate how stable and repeatable these measures are at the individual level across years; and (3) model whether any demographic and life-history variables (age and sex) are related to this variation. We provide future directions for the wider roll-out of the method and suggestions for next steps of data collection and analysis.

2 | MATERIALS AND METHODS

2.1 | Study site and subjects

Data were collected from the wild chimpanzee field research site of Bossou, southeastern Guinea, West Africa (Humle, 2011b). Long-term observations of a single community of wild chimpanzees *Pan troglodytes verus* have been conducted here since 1976 (Matsuzawa et al., 2011; Sugiyama, 1984). In 1988, researchers established a quasi-experimental 'outdoor laboratory' in a natural forest clearing (Matsuzawa, 1994) to observe nut cracking behaviour, which peaks naturally in June–July and November–December (Yamakoshi, 1998). In this study, data were collected across 17 years (1996–2017) from film recording at the outdoor laboratory during each dry season (December–February) whereby provisioning of nuts and stones by researchers allowed effective observation of nut-cracking behaviour. Approximately 20–30 h of footage was collected annually using tripod-mounted human-operated cameras, sometimes from multiple angles. The outdoor laboratory was opened daily during daylight hours throughout the field season (typically across 1–2 months), and video was only recorded if individuals attended, until they left the forest clearing. Video data collection was non-invasive and approved by the ethics committee at Kyoto University's Primate Research Institute. All field research in Guinea conformed to ethical guidelines outlined by the Association for the Study of Animal Behaviour.

2.2 | Video pre-processing

To minimise computation for image processing and data storage, while ensuring consistent sampling across years, an average of 6 h of video content (5.99 ± 1.46 h) distributed across 15.29 ± 24.14 videos was randomly sampled per field season for 17 years spanning 1996–2017, totalling 101.82 h of footage from manually digitised tapes (see Table S1). To avoid data re-sampling, maintain consistency and improve performance of our recognition model, we used one medium-distance camera angle per year to capture clear footage of individual faces, avoiding the low resolution from wider camera angles where face detection is challenging. Videos were processed using a deep learning face recognition pipeline as described in Schofield et al. (2019) using Pytorch (Paszke et al., 2019) and a 'Titan X' Graphical Processing Unit (GPU). For each video in the dataset, a CSV file produced a spatiotemporal database with coordinates of individual detections in each video frame. Multiple CSV files were collated into a large dataframe comprising 7 million frame-level detections using the package PANDAS (McKinney, 2011) in Python version 3.7.3 (Oliphant, 2007).

The face recognition method uses supervised deep neural networks and makes a prediction for each face detection from 23 fixed

identity classes (and an additional class for mistaken 'nonface' false-positive detections). Some individuals were rarely sighted and thus not included in training data (2002 $n=1$, 1996–1999 $n=1$). For these 'unknown' chimpanzees, the model is forced to falsely predict an identity from one of the fixed classes; to reduce the number of false positives, individuals known not to be present in that year (those yet to be born, or having died or dispersed) were removed from the dataframe using field notes/records. Continuous video sequences often contain shot-changes—jumps in time where the camera was switched on and off. To prevent an impact on the social network analysis (individuals being incorrectly grouped/associated across different time periods), a Python shot-detector (Castellano, 2021) was used with its default sensitivity value to mark these points. Given the variable frame rate of videos, frames were converted into seconds for aggregation to reduce the amount of observations for analysis and provide consistent units of time for sampling networks.

2.3 | Social network adjacency matrices

To quantify social networks of the Bossou group, we adopted the 'Gambit of the group' methodology, which calculates association scores based on the observation of all individuals within a group at a single point in time, rather than using behavioural interactions such as grooming or physical contact (Franks et al., 2010). Videos are split into a set of fixed time windows/observation points (or 'grouping events') where individuals detected in the same time window are recorded as co-occurring in that group. This produces networks represented as weighted adjacency matrices, where the association scores for each pair of individuals (dyads) are weighted using the simple ratio index (Hoppitt & Farine, 2018; see section below) to reflect the proportion of times they co-occurred at the outdoor laboratory (Schofield et al., 2019). These matrices form the basis of social networks, where 'nodes' represent individual chimpanzees, and 'edges' the strength of association between individuals. These association-based networks thus infer social connections without measuring them directly, but when sampled at high rates produce robust association indices for easily observed groups (Farine & Whitehead, 2015; Silk et al., 2015).

2.4 | Correlations across aggregation rates

In the context of video data collection at the outdoor laboratory, individuals were recorded at approximately a 15–20m distance. The general protocol for the camera shots was to capture all individuals within the video frame, covering most of the 7×20m space of the outdoor laboratory. However, occasionally, the camera would pan or zoom to capture specific bouts of activity, such as a special case of tool use or social interaction, which may only include the 1–5m surrounding the focal individuals (Figure 1). Wider shots are more likely to capture all individuals at the forest clearing, whereas zoom shots capture individuals in close proximity. The size of the
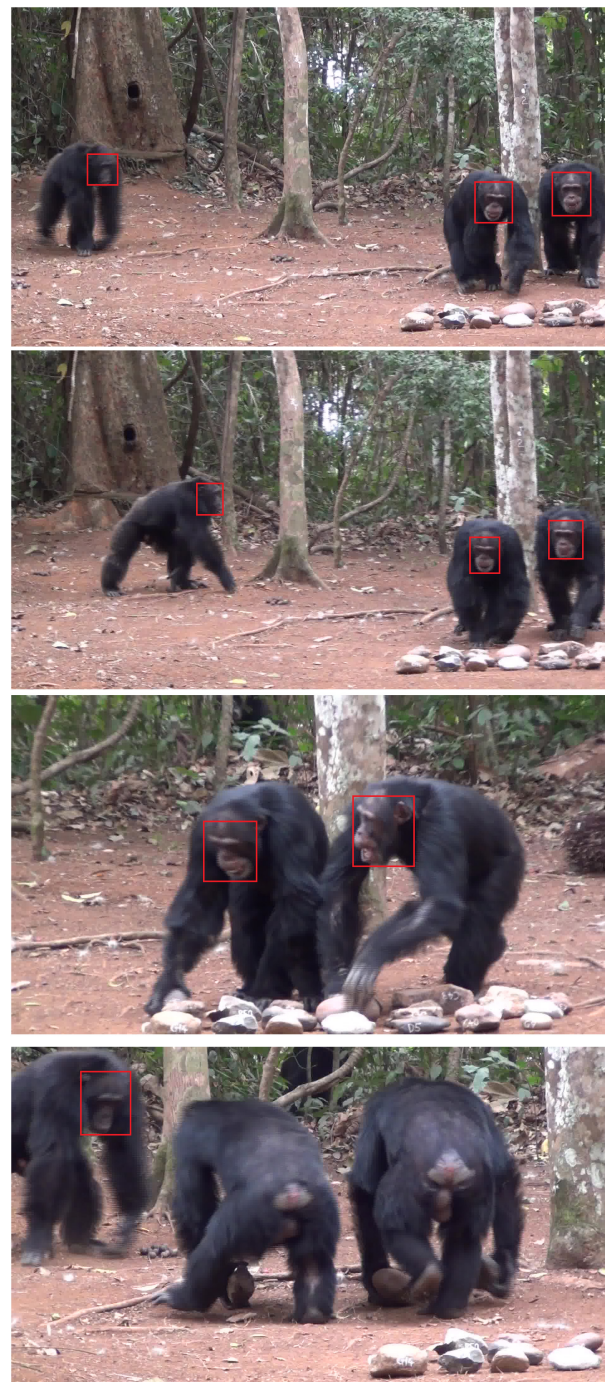
FIGURE 1 Sequence from a 15-s clip of footage filmed at the outdoor laboratory in 2012. The example detections depict three adult males and highlight shot variations resulting from the moving camera, as well as potential biases caused by missed detections when individuals turn their heads or move out of shot. Images 1–4 (top to bottom) show (1–2) a medium shot of three individuals entering; (3) a zoom shot of two males (with one peripheral male out of shot); and (4) two male turning their backs and being missed by our detector.

temporal window used for recording social associations thus has key implications for capturing social associations. Sometimes, the detector can fail to detect faces when they are obscured from view (e.g. by vegetation or another individual) or the individual moves

out of shot (e.g. Figure 1, 3–4). Larger time windows (e.g. spanning multiple seconds, minutes or even hours) may be less sensitive to these detection errors, as only a single detection is needed for individuals to co-occur in the time window—this means individuals can co-occur, despite having moved temporarily out of shot (the observational area). Larger time windows may group together individuals who never interacted—for example, if they enter and leave the forest clearing at different points in time. Furthermore, a single misidentification in an image can incorrectly aggregate individuals across this larger time range (e.g. a 10-min interval).

To examine how temporal window size impacts gregariousness scores and social networks, we constructed weighted adjacency matrices for a broad range of temporal windows/fixed aggregation rates (1–3600 s at 10 s intervals), resulting in 360 methodological replicates. Social networks were generated and analysed in R v. 4.1.0 (R Core Team, 2021). For each aggregation rate, we generated a 'group-by-individual' matrix detailing the co-occurrences between all chimpanzees, which was weighted using the simple ratio index (SRI): $\text{Detections}_{A,B} / (\text{Detections}_A + \text{Detections}_B - \text{Detections}_{A,B})$ (Cairns & Schwager, 1987). This represents the proportion of grouping events/time windows in which focal individuals were detected together. One social network was produced for each field season, which was resampled using each aggregation rate, producing 6120 networks for comparison for 17 years of data. To examine how varying the size of the temporal window affected the sociality scores, we used Pearson's correlation coefficient to compare social network values (strength) calculated from weighted networks generated using face recognition ('AI') at different aggregation rates (1–3600 s) against the most extreme aggregation rate (every second). We also mapped how the mean values for Time Alone, Mean Group Size and Network Density changed as a function of aggregation rate across this time interval range.

2.5 | Statistical analysis

2.5.1 | Sociality measures

We generated common individual-level network and sociality variables used to study animal social networks (Farine & Whitehead, 2015) and gregariousness across the lifespan in primates (Machanda & Rosati, 2020). Individual metrics were used to examine correlations between different sampling methods, and as response variables in Generalised Linear Models (GLMMs). We calculated measures including (1) 'Strength'—the sum of an individual's social associations or the proportion of time spent associating with others; (2) 'Mean Group Size'—the average number of co-detections at any one time across all time windows as a measure of party size and (3) 'Time Alone'—the proportion of 'grouping events'/time windows where an individual was detected alone as a gauge of solitariness. We calculated a 'global' network measure to quantify a feature of the overall structure of the network ('Density')—the proportion of edges (social connections) to the total potential edges in the

weighted network—to examine how aggregation rates affected the overall cohesiveness and connectivity of the network.

2.5.2 | Models

For every year in the dataset, each individual has one value for each of our outcome variables/gregariousness scores (conditional on being detected by our recognition model at least once in that year). Consistent experimental and filming protocols at a single location and during the same time of year reduced the potential impact of nuisance/extraneous variables on our sociality scores (e.g. such as from variations caused by environmental factors, locality and seasonality). For two variables (Strength and Mean Group Size), we fitted linear mixed models (LMMs) with Gaussian error structure and identity link function using the package LME4 (Bates et al., 2007) in the R v. 4.1.0 statistical SOFTWARE package (R Core Team, 2021). To model the proportion data for 'Time Alone', we fitted a GLMM with a beta regression error structure, logit link function and Restricted Maximum Likelihood (REML) using the package GLMMTMB (Brooks et al., 2017). To fit the beta regression model using glmmTMB, the zero-inflated proportion scores (individuals who were not observed alone) were adjusted by 0.001. We examined whether demographic variables (Age and Sex) influenced our three individual sociality metrics (sampled at an intermediate 120 s aggregation rate). All model sets retained the same structure for fixed and random effects: Individual Identity and Year were included as categorical random effects; Sex, Age in years, and Population Size were included as fixed effects. Given that age has been shown to follow a quadratic curve in primate sociality and cognition (Lacreuse et al., 2020), we included it as a quadratic term for all models. We scaled all of our continuous predictor variables to have a mean of 0 and a standard deviation of 1 to encourage model fitting and aid interpretability (Schielzeth, 2010). No further transformation was applied to the quadratic age term (which represents the square of the z-transformed Age variable). To examine the effect that our test variables (Age and Sex) had on individual sociality scores, we used likelihood ratio tests (Harrison et al., 2018; Nakagawa & Schielzeth, 2010) comparing our full model containing the fixed predictors of interest, and a reduced regression/'null' model, which only contained our control predictor variables (Year of Observation and Individual as random effects, and Population Size as a control fixed effect). This approach differs from the use of null models that rely on randomisations or permutations to address bias in network studies (e.g. Farine & Carter, 2022). For each variable, a further likelihood test was performed comparing the full model with a null model with individual identity as a random effect removed to examine the impact it had on our outcome variables (Nakagawa & Schielzeth, 2010; see Section 2.5.3 below). For each full-null model comparison, we used the 'anova' and 'compare' functions in the PERFORMANCE package in R (Lüdecke et al., 2021) to assess model fit and obtain *p*-values. For all models, we checked for collinearity, homogeneity of variance and the normality of residuals using the 'check_model' function in the PERFORMANCE package, which did

not reveal any significant abnormalities. 'Time Alone' was modelled using beta regression which showed slight violation of normality of residuals due to outliers in the dataset, indicating that our model under-predicted some large values at the tail end of the distribution (see Figure S6 for normality checks and discussion of outliers; Vallejo Seco et al., 2013).

2.5.3 | Repeatability

Repeatability measures the consistency of inter-individual differences over time. Traits that show a high level of repeatability have high between-individual variance and low within-individual variance, which indicates that individuals have differing and stable tendencies or strategies for interacting/associating with group members over time (Tkaczynski et al., 2020). To identify the individual-level repeatability of these metrics, we assessed the proportion of variance in our gregariousness variables that were attributable to the individual identity random effect in our models (Nakagawa & Schielzeth, 2010). We compared full models containing all fixed and random effects, and null models without the random intercept and slope for individual identity, to test whether individual identity explained variance in our outcome variables and improved model fit (Tkaczynski et al., 2020). We used the 'r.squaredGLMM' function from the MuMin package in R (Bartoń, 2021) to calculate individual R^2 for a model containing individual identity as the only random effect (dropping Year) to examine the repeatability/strength of the effect that individual identity had.

3 | RESULTS

3.1 | Aggregation rates

Figure 2 summarises the changes in correlations and mean values of social network and sociality variables across rates aggregation. For Strength (Figure 2a), we found a steep drop-off in correlation values when comparing networks from the highest aggregation rate (1 s) to other aggregation rates, indicating that the highest and lowest aggregation rates produced divergent social network scores. This indicates that extreme aggregation rates may be capturing different association patterns, and may also be introducing errors affecting the network scores. The network density ranged from 0.94 to 0.99 (Figure 2c) which indicates that as the size of the aggregation time window increases, a greater proportion of individuals are observed together, reducing variation in association rates and network scores. This high range of connectivity in the network is also likely due to the small group size from population decline (Humle, 2011a), and a feature specific to Bossou (Sugiyama, 1988, 2004; Sugiyama & Fujita, 2011) compared to other populations (e.g. Langergraber et al., 2009; Mitani, 2009) (see discussion in Supporting Information Section 'Social and ecological context at Bossou'). Sociality variables ('Mean Group Size' and 'Time Alone') also appear to be sensitive to

the size of the temporal window, with the highest aggregation rate producing more extreme values before stabilising at around the 600 s aggregation rate. This suggests that short temporal windows (below a minute) are susceptible to bias when the camera zooms or captures a single individual in shot, therefore overestimating the time individuals spend alone and underestimating the party size. Using a low aggregation rate with larger time windows has the inverse effect, reducing the number of observations, overestimating party size and underestimating individual solitariness as more individuals are grouped together. Collectively across measures, visual assessment of these curves indicates that an intermediate value of 50–120 s avoids extreme values and the potential noise generated from oversampling, whilst still capturing enough variation in the networks to make inferences about individual-level sociality patterns; however, this optimal window may shift depending on the time span of the study and amount of data collected.

3.2 | Modelling results

Effect sizes and measures of model fit from likelihood ratio tests for each model set are presented in Tables 1 and 2. We found no significant differences between the effect of age and sex on sociality scores for Strength ($\chi^2 = 5.18$, df = 2, $p = 0.075$) and Mean Group Size ($\chi^2 = 2.58$, df = 2, $p = 0.27$). We found weak evidence for a quadratic age effect on Time Alone (AIC difference = 2.47; Figure S8). Repeatability scores for each sociality variable are presented in Table 3. For all models, individual identity improved model fit: Strength ($\chi^2 = 116.14$, df = 1, $p < 0.001$), Mean Group Size ($\chi^2 = 11.70$, df = 1, $p < 0.001$) and Time Alone ($\chi^2 = 20.11$, df = 1, $p < 0.001$). This indicates that Bossou chimpanzees showed inter-individual variation and high repeatability (consistency) in gregariousness over time. For all of our models, individual identity had a significant impact on model fit; however, the size of the effect varied—the variation explained by the random effect of individual identity was highest for Strength (0.35) followed by Time Alone (0.13) and Mean Group Size (0.03). Inter-individual variation in Strength, Mean Group Size and Time Alone are visualised in Figure 3.

4 | DISCUSSION

In this study, we tested the reliability of automated face recognition for measuring association-based social networks and individual-level gregariousness scores, using a long-term video archive dataset of wild chimpanzees. We examined how the size of the temporal window used for capturing co-occurrences affects the robustness of estimates of chimpanzee association scores over time, and the impact of potential errors and biases generated from different rates of aggregation on these measures. Overall, our results reveal that the size of temporal window used for aggregation of social networks should be a key consideration when measuring gregariousness using computer vision-based identification of individuals in video footage.

FIGURE 2 Networks aggregated using different fixed time windows/aggregation rates (1–3600 s). (a) Strength values from each network generated at different aggregation rates, correlated using Pearson's to the smallest temporal window (1 s). (b) Mean value for group size and Time Alone (scaled to be between 0 and 1) for each aggregation rate. (c) Network density for each network and aggregation rate. Dashed lines indicate range for intermediate aggregation (50–120 s) which avoids extreme values.
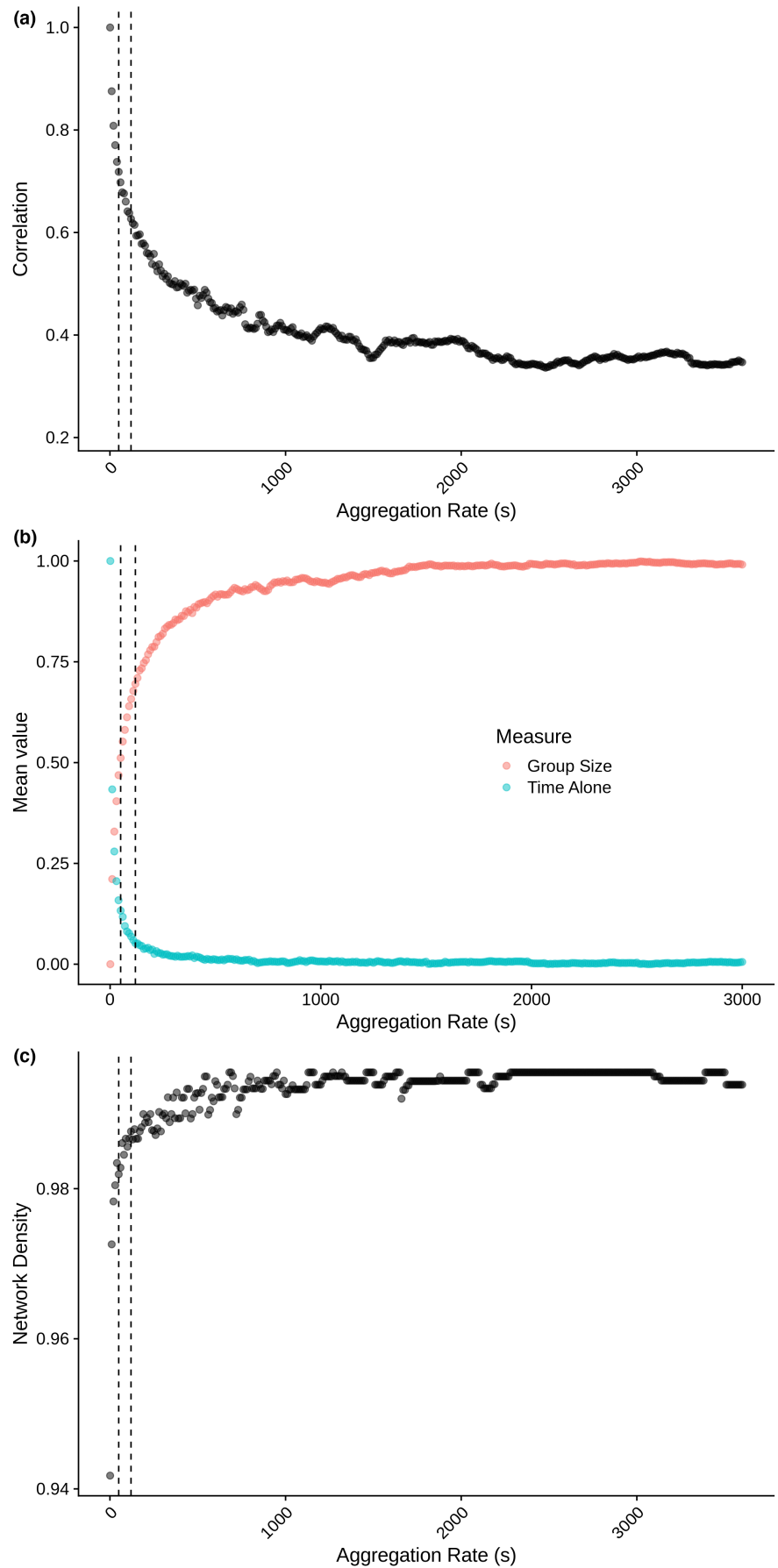
Although networks produced across different aggregation rates were generally well correlated, extreme aggregation rates produced divergent scores, and our sociality variables Mean Group Size and Time Alone showed a steep drop off in values (Figure 2b), indicating that care should be taken to evaluate how variables change as a function of aggregation rate. We found that selecting an intermediate aggregation rate avoids potential noise and extreme values generated from under- and over-aggregating. This approach improves the researcher's ability maximise the intensity of data collection while reducing errors, to capture meaningful variation in networks and gregariousness scores. However, the behavioural context of the study species and the choice of measures should inform the sampling strategy and size of temporal window used for capturing social behaviour.

TABLE 1 Model results: Overview of model parameters for linear mixed models fitted using the LMER and GLMMTMB packages in R.

Outcome variable	Term	Estimate	SE
Strength	Intercept	3.50	0.35
	Sex	−0.93	0.47
	Age^2	−0.11	0.08
	Population size	0.79	0.17
Mean Group Size	Intercept	3.50	0.35
	Sex	−0.93	0.47
	Age^2	−0.11	0.08
	Population size	0.79	0.17
Time Alone	Intercept	−2.91	0.18
	Sex	−0.24	0.24
	Age^2	−0.20	0.08
	Population size	−0.29	0.08

4.1 | Robustness of method

Our dataset was sampled annually within a 1–2 month window of data collection spanning 17 years. This is within the range of observation time for other studies generating animal social networks from intensive data collection over short periods (Davis et al., 2018; Feczko et al., 2015; Gelardi et al., 2020). The correlation of networks produced using larger temporal windows (lower aggregation rates) appeared to break down at around 1000s (Figure 2a), indicating that networks and measures generated at this aggregation rate were unreliable, which may be driven in part by too few data points and the sparse networks it produced. Examining the performance of the models, individual repeatability and the distributions of sociality variables across all aggregation rates could help to investigate the source of diverging values and potential bias in the networks further. For other automated tracking technologies, such as GPS, there is often a trade-off between rate of sampling and the longevity of battery life (He et al., 2022; McCann et al., 2021); sparse sampling can be surprisingly low for producing robust association networks if collected over many months (Davis et al., 2018). Thus, consideration should be given to the span of data collection in video databases—for our small seasonal snapshots, more hours of observation within each year would be required to generate robust measures using larger temporal windows. Data collected consistently across the field season through continuous monitoring (such as remote camera traps) may generate robust and stable network estimates using larger windows of aggregation. In the context of our study, using high aggregation rates may more closely approximate proximity networks (measuring how often individuals are within several metres of one another), whereas sparser aggregation rates may approximate party membership at a coarser level. In our study, we have also not examined how sparse networks from using larger temporal windows produce false positives, due to the potential impact of a single misidentification for aggregating individuals within a large time

TABLE 2 Likelihood ratio test comparing model fit between full model containing all fixed and random effects, and a null model with fixed predictor variables of interest (age and sex) removed for each of our outcome variables: 'Strength', 'Mean Group Size' and 'Time Alone'. For each likelihood ratio test, we report the AIC, BIC and R^2 for the full and null model, and the difference between them (null-full) when the predictors of interest (age and sex) are removed from the full model. We also present χ^2 and p-value for significance, and AIC for glmmTMB beta regression model (a difference of ~2 or less in AIC indicates no difference between models).

Outcome variable	Model comparison	AIC	BIC	Conditional R^2	Marginal R^2	χ^2	df	p value
Strength	Full model	537.61	561.62	0.88	0.33	—	—	—
	Null model	538.79	555.94	—	—	—	—	—
	Full-null	1.18	−5.68	—	—	5.18	2	0.075
Mean Group Size	Full model	657.48	681.49	0.78	0.52	—	—	—
	Null model	656.06	673.21	—	—	—	—	—
	Full-null	−1.42	−8.28	—	—	2.58	2	0.27
Time Alone	Full model	−1027.52	−1003.52	0.27	0.08	—	—	—
	Null model	−1030.00	−1012.85	—	—	—	—	—
	Full-null	−2.47	−9.33	—	—	—	—	—

TABLE 3 Individual chimpanzee repeatability in sociality variables. Variance in outcome variables attributed to individual identity as a random effect in linear models (R^2), spanning 17 years of video footage, and aggregated at 120 s time windows. We report the AIC/BIC values for the full/null model with the random effect of identity removed, and the χ^2 and the p-value significance of this difference. We report 'ID model', which is the full model with ID as the only random effect (dropping year), to show the strength of the effect of individual ID, as a measure of individual-level repeatability of each trait. Individual R^2 is conditional R^2 (variance of fixed and random effects combined)—R^2 (the variance of only fixed effects).

Outcome variable	Model comparison	AIC	BIC	Conditional R^2	Marginal R^2	χ^2	df	*p* value
Strength	Full model	537.61	561.62	0.88	0.33	—	—	—
	Null model	651.75	672.32	—	—	—	—	—
	Full-null	114.14	110.71	—	—	116.14	1	<0.001
	ID model	—	—	0.71	0.35	0.35	—	—
Mean Group Size	Full model	657.48	681.49	—	—	—	—	—
	Null model	667.19	687.76	—	—	—	—	—
	Full-null	9.70	6.28	—	—	11.70	1	<0.001
	ID model	—	—	0.58	0.56	0.03	—	—
Time Alone	Full model	1027.52	1003.52	—	—	—	—	—
	Null model	1009.41	988.84	—	—	—	—	—
	Full-null	18.11	14.68	—	—	20.11	1	<0.001
	ID model	—	—	0.23	0.10	0.13	—	—

window, and in future additional thresholds can be implemented to reduce these errors (e.g. requiring a certain amount of time for individuals to spend together to co-occur within a time window).

Our results indicate that our recognition model produced individual-level estimates of sociality that were replicable over time at our intermediate rate of aggregation, highlighting the robustness of the method to produce reliable sociality estimates. Our finding that gregariousness was variable between individuals, and consistent within individuals across years, is consistent with the wider pattern shown by many other species, namely that behavioural patterns such as social network positions remain stable across the lifespan (Aplin, Firth, et al., 2015; Blaszczyk, 2017; Finger et al., 2017; Jacoby et al., 2014; Wuerz & Krüger, 2015). This also supports previous findings on repeatability of sociality in other chimpanzee groups which indicate that gregariousness becomes fixed during development through some form of canalisation, and remains consistent throughout their lifetime (Altschul et al., 2018; Tkaczynski et al., 2020).

4.2 | Extending to other species and contexts

This system was developed and tested for a single group of chimpanzees, and thresholds are likely to vary depending on the ecological context and species (Webster & Rutz, 2020). However, the methodology for model training is not species specific (Schofield et al., 2019), and it is relatively simple to implement as it only requires a time series of individual detections, which means it can be flexibly applied for different forms of video data collection, species and ecological contexts. Our approach using Gambit of the Group for co-occurrences was able to generate enough variation and

consistency in social measures, despite the relatively low image quality of archival data, moving cameras, small population size and specific context of a single forest clearing at Bossou. For video datasets collected over short timeframes (e.g. over several hours, days or weeks), in constrained contexts where many individuals can be observed clearly in the same shot or do not have freedom to vary their social groups (e.g. in an enclosure in captivity), then additional classifiers which track proximity or behavioural interactions may be required (see future directions below). The approach outlined here is particularly useful for species that exhibit fission–fusion dynamics and for data collection where grouping events are recorded over many months (Franks et al., 2010; Silk et al., 2015) and thus can be applied for either hand-held video focal follow data or long-term camera trap projects. For any approach, consideration of the size and quality of the shot should be considered, and studies using automated recognition classifiers should maximise the field of vision and image quality to capture all visible individuals whilst maintaining a close enough shot to detect individual faces reliably. The use of a full body recognition classifier should be a consideration due to its potentially greater recall than face recognition, depending on the size of the species, how recognisable individuals are by their bodies and the ecological context (e.g. how often individuals are occluded by vegetation; Bain et al., 2019; Ferreira, Silva, et al., 2020).

4.3 | Benefits of using AI for animal sociality

A significant advantage of this system is that once the neural networks have been trained, archive footage can be used to optimise strategies for data collection of social behaviour from longitudinal
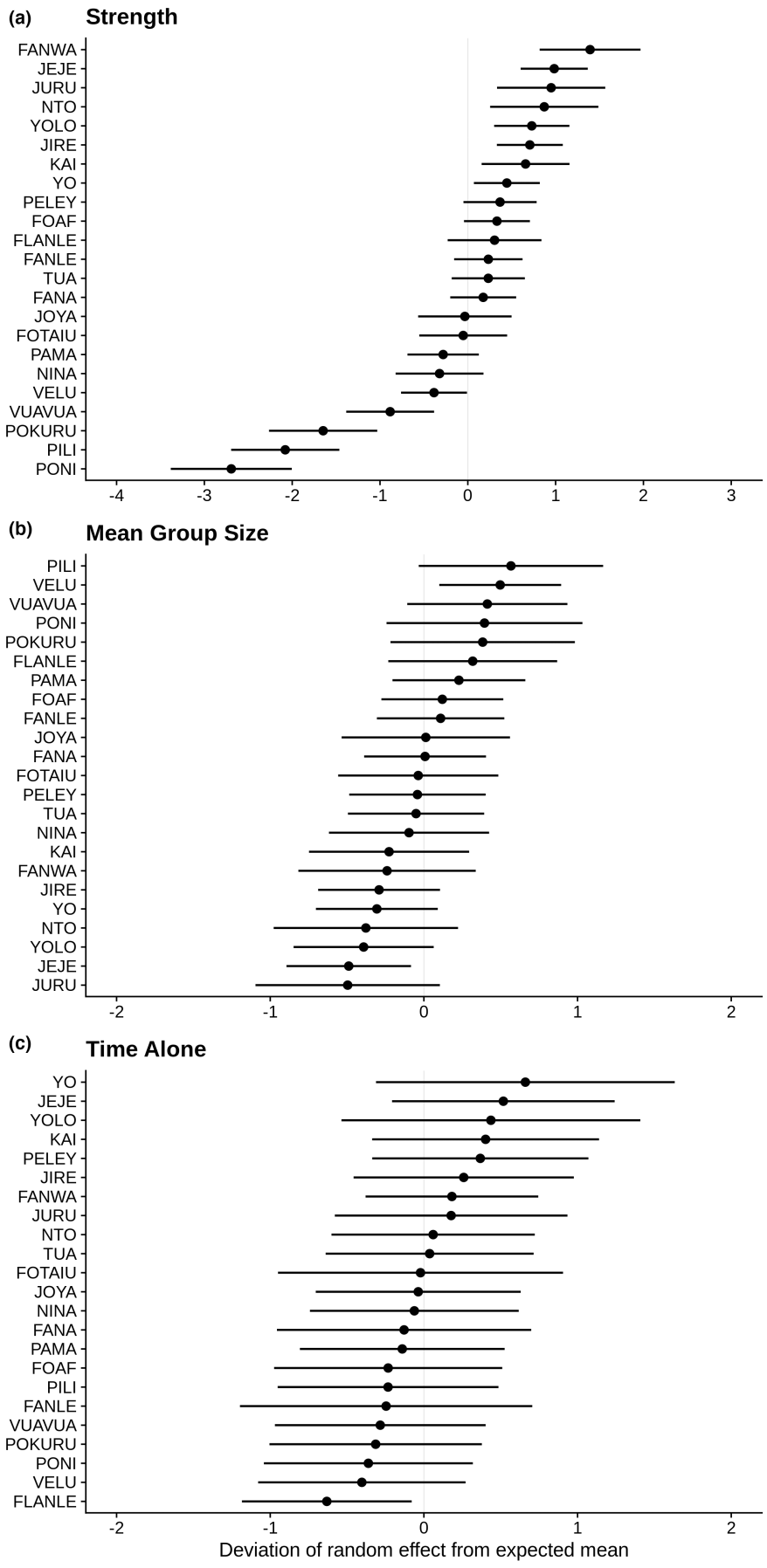
FIGURE 3 Individual random effect coefficients (*x*-axis indicates mean and variation) representing inter-individual variation in chimpanzee sociality measures for Strength, Mean Group Size and Time Alone at the 120s aggregation rate. Scores above or below zero represent higher or lower than average expression in outcome variables for each individual. Error bars indicate 95% confidence intervals for the effects.

datasets. Manual data collection of social networks undertaken retrospectively via video coding or through behavioural observation in situ is time-consuming and can take many times the length of the video depending on the task (Anderson & Perona, 2014). Our method for generating social indices is entirely automatic, so once models are trained no video coding is required, which reduces potential for human observer bias and improves scalability; previous work found that without contextual information researchers struggle to identify individuals from video footage, being outperformed by the recognition model in both speed and accuracy for this fine level task (Schofield et al., 2019). This is particularly relevant for long-term field sites where researchers with different levels of experience and interest can introduce variation in how data are collected from year to year. A key problem in the field of animal social networks is generating sufficient data for a robust analysis (Farine & Whitehead, 2015; Franks et al., 2010). This automated approach is scalable and can increase the quantity of observations, increasing the statistical power and reliability estimation of the 'real' networks, without the need for re-watching and coding video.

4.4 | Future directions

Despite these significant advantages, there are limitations of using AI. Deep neural networks take significant time and resources in terms of storage, annotation and processing of video data. Currently, investment is required by engineers to build, maintain and run the pipelines; these require technical knowledge, reducing the availability and practicality of the approach in the field and leading to bottlenecks for data preparation, annotation, storage and training between engineers and biologists. In order for this technology to be more widely adopted, open datasets and end-to-end pipelines need to become more transparent, accessible and efficient to use through open-access software projects (Mathis & Mathis, 2020; Nath et al., 2019). The original development for our system took over a year to implement; however, the bulk of this time was spent on acquisition of the data, expert annotation and development of the tools to train models from scratch. However, it should be possible to train and fine-tune models in much less time as more open frameworks and models are released and computational resources become more affordable. Supervised methods are useful for retrospective analysis of complete archives when individuals are clearly visible and identifiable, as is the case with our dataset. However, applying these methods becomes significantly more challenging for large unhabituated populations with unknown individuals. Annotating data for training requires significant effort, and incorporating new individuals into the model may require modifying the final layer of the neural network and retraining (Schofield et al., 2019). New methods are emerging to improve the efficiency and flexibility of computer vision pipelines. Active learning models can automatically locate images of interest for human annotators, and lightweight on-device/real-time recognition systems can improve the ease with which data are stored and annotated (Norouzzadeh et al., 2021). New methods for

re-identification of individuals are emerging (Miele et al., 2021), such as metric learning, which can quantify similarity between images in a dataset of individuals, or one-shot meta-learning, training networks to learn categories without human annotation, without needing to retrain from scratch, or when classes are data poor (Hospedales et al., 2021).

The flexibility of such systems will enable development of more sophisticated analysis—for example, layering individual identification with behavioural recognition (Bain et al., 2021). Such new approaches will enable capture of a richer array of animal behaviour, to examine how individual-sociality, behaviour and networks interact in real time—for example, to trace how new behaviours spread through a population via social learning, using dynamic network-based diffusion analysis (Hasenjager et al., 2021; Hobaiter et al., 2014). Despite over 60 years of field research, few studies have leveraged a longitudinal approach to examine how sociality patterns change over time and are impacted by demographics, such as reductions in population size, the loss of key individuals to death or dispersal, lack of emigration, reductions in fertility and ageing population structures (Brent et al., 2017; Machanda & Rosati, 2020; Rosati et al., 2020; Sheldon et al., 2022; Shizuka & Johnson, 2019) (see discussion in Supporting Information Section 'Sociality over the lifespan'). For large-camera trap databases and continuous monitoring projects, systematic comparison of multiple communities using automated classifiers for video data has potential to examine behavioural variation of many animal groups across large temporal and geographical ranges.

5 | CONCLUSIONS

Capturing social associations using automated face recognition and social networks is a powerful tool for unlocking video archives and revealing the social structure of wild animal populations; however, investigation of the adequate thresholds for defining edges is required to provide robust insights into individual-level patterns of sociality. Overall, we find that automated face recognition using deep neural networks provides a robust method for producing chimpanzee social networks and sociality measures, and brings several key advantages including reduction in time investment in data collection and human observer bias, high-throughput sampling for statistical power, and optimisation of network aggregation strategy for modelling biological processes. The method produces repeatable social networks, which show that inter-individual differences in sociality patterns remain stable over time in a group of wild chimpanzees. Using this automatic framework, we find that although there was individual-level variation in social behaviour, age and sex do not appear to be strong drivers of individual gregariousness and social structure for the Bossou community in this data. This work provides a framework for evaluating the consistency and reliability of using AI to measure social behaviour from longitudinal archive datasets; it moves towards the widespread adoption of numerous deep learning algorithms which are able to generalise behavioural data collection across multiple

species and habitats. Ultimately, this methodology builds the platform to model the flexibility and variation in social behaviour and social networks in different ecological contexts and populations, over larger temporal and geographic ranges and across the lifespan in wild animals.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14181.

DATA AVAILABILITY STATEMENT

Data and code used in this study are available in a Github repository at https://github.com/dschof/DeepPrimateNetworksMEE and archived in Zenodo under https://zenodo.org/record/8063792 (Schofield et al., 2023).

ORCID

Daniel P. Schofield 🄳 https://orcid.org/0000-0002-3308-0209
Gregory F. Albery 🄳 https://orcid.org/0000-0001-6260-2662
Josh A. Firth 🄳 https://orcid.org/0000-0001-7183-4115
Alexander Mielke 🄳 https://orcid.org/0000-0002-8847-6665
Misato Hayashi 🄳 https://orcid.org/0000-0001-7289-6414
Tetsuro Matsuzawa 🄳 https://orcid.org/0000-0002-8147-2725
Dora Biro 🄳 https://orcid.org/0000-0002-3408-6274
Susana Carvalho 🄳 https://orcid.org/0000-0003-4542-3720

REFERENCES

Abbot, P., & Rubenstein, D. (2017). *Comparative social evolution*. Cambridge University Press.

Alexander, R. D. (1974). The evolution of social behavior. *Annual Review of Ecology and Systematics, 5*(1), 325–383.

Allen, J., Weinrich, M., Hoppitt, W., & Rendell, L. (2013). Network-based diffusion analysis reveals cultural transmission of lobtail feeding in humpback whales. *Science, 340*(6131), 485–488.

Altschul, D. M., Hopkins, W. D., Herrelko, E. S., Inoue-Murayama, M., Matsuzawa, T., King, J. E., Ross, S. R., & Weiss, A. (2018). Personality links with lifespan in chimpanzees. *eLife, 7*, e33781.

Anderson, D. J., & Perona, P. (2014). Toward a science of computational ethology. *Neuron, 84*(1), 18–31.

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature, 518*(7540), 538–541.

Aplin, L. M., Firth, J. A., Farine, D. R., Voelkl, B., Crates, R. A., Culina, A., Garroway, C. J., Hinde, C. A., Kidd, L. R., Psorakis, I., Milligan, N. D., Radersma, R., Verhelst, B. L., & Sheldon, B. C. (2015). Consistent individual differences in the social phenotypes of wild great tits, parus major. *Animal Behaviour, 108*, 117–127.

Bain, M., Nagrani, A., Schofield, D., Berdugo, S., Bessa, J., Owen, J., Hockings, K. J., Matsuzawa, T., Hayashi, M., Biro, D., Carvalho, S., & Zisserman, A. (2021). Automated audiovisual behavior recognition in wild primates. *Science Advances, 7*(46), eabi4883.

Bain, M., Nagrani, A., Schofield, D., & Zisserman, A. (2019). Count, crop and recognise: Fine-grained recognition in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*.

Bartoń, K. (2021). MuMIn: Multi-model inference. https://cran.r-project.org/package=mumin.

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. R package version 2(1), 74.

Blaszczyk, M. B. (2017). Consistency in social network position over changing environments in a seasonally breeding primate. *Behavioral Ecology and Sociobiology, 72*(1), 11.

Brent, L. J. N., Ruiz-Lambides, A., & Platt, M. L. (2017). Family network size and survival across the lifespan of female macaques. *Proceedings of the Biological Sciences, 284*(1854), 20170515.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400.

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology, 52*(3), 675–685.

Cairns, S. J., & Schwager, S. J. (1987). A comparison of association indices. *Animal Behaviour, 35*(5), 1454–1469.

Castellano, B. (2021). Pyscenedetect. https://github.com/breakthrough/pyscenedetect.

Castles, M., Heinsohn, R., Marshall, H. H., Lee, A. E. G., Cowlishaw, G., & Carter, A. J. (2014). Social networks created with different techniques are not comparable. *Animal Behaviour, 96*, 59–67.

Chen, P., Swarup, P., Matkowski, W. M., Kong, A. W. K., Han, S., Zhang, Z., & Rong, H. (2020). A study on giant panda recognition based

on images of a large proportion of captive pandas. *Ecology and Evolution*, 10(7), 3561–3573.

Clutton-Brock, T., & Sheldon, B. C. (2010). Individuals and populations: The role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, 25(10), 562–573.

Croft, D. P., James, R., & Krause, J. (2008). *Exploring animal social networks*. Princeton University Press.

Davis, G. H., Crofoot, M. C., & Farine, D. R. (2018). Estimating the robustness and uncertainty of animal social networks using different observational methods. *Animal Behaviour*, 141, 29–44.

Dore, K. M., Hansen, M. F., Klegarth, A. R., Fichtel, C., Koch, F., Springer, A., Kappeler, P., Parga, J. A., Humle, T., Colin, C., Raballand, E., Huang, Z.-P., Qi, X.-G., Di Fiore, A., Link, A., Stevenson, P. R., Stark, D. J., Tan, N., Gallagher, C. A., … Fuentes, A. (2020). Review of GPS collar deployments and performance on nonhuman primates. *Primates*, 61(3), 373–387.

Farine, D. R. (2015). Proximity as a proxy for interactions: Issues of scale in social network analysis. *Animal Behaviour*, 104(104), e1–e5.

Farine, D. R. (2018). When to choose dynamic vs. static social network analysis. *Journal of Animal Ecology*, 87(1), 128–138.

Farine, D. R., & Carter, G. G. (2022). Permutation tests for hypothesis testing with animal social network data: Problems and potential solutions. *Methods in Ecology and Evolution*, 13(1), 144–156.

Farine, D. R., & Whitehead, H. (2015). Constructing, conducting and interpreting animal social network analysis. *The Journal of Animal Ecology*, 84(5), 1144–1163.

Feczko, E., Mitchell, T. A. J., Walum, H., Brooks, J. M., Heitz, T. R., Young, L. J., & Parr, L. A. (2015). Establishing the reliability of rhesus macaque social network assessment from video observations. *Animal Behaviour*, 107, 115–123.

Ferreira, A. C., Covas, R., Silva, L. R., Esteves, S. C., Duarte, I. F., Fortuna, R., Theron, F., Doutrelant, C., & Farine, D. R. (2020). How to make methodological decisions when inferring social networks. *Ecology and Evolution*, 10(17), 9132–9143.

Ferreira, A. C., Silva, L. R., Renna, F., Brandl, H. B., Renoult, J. P., Farine, D. R., Covas, R., & Doutrelant, C. (2020). Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*, 11(9), 1072–1085.

Finger, J. S., Guttridge, T. L., Wilson, A. D. M., Gruber, S. H., & Krause, J. (2017). Are some sharks more social than others? Short- and long-term consistencies in the social behavior of juvenile lemon sharks. *Behavioral Ecology and Sociobiology*, 72(1), 17.

Firth, J. A., & Sheldon, B. C. (2016). Social carry-over effects underpin trans-seasonally linked structure in a wild bird population. *Ecology Letters*, 19(11), 1324–1332.

Franks, D. W., Ruxton, G. D., & James, R. (2010). Sampling animal association networks with the gambit of the group. *Behavioral Ecology and Sociobiology*, 64(3), 493–503.

Gelardi, V., Godard, J., Paleressompoulle, D., Claidiere, N., & Barrat, A. (2020). Measuring social networks in primates: Wearable sensors versus direct observations. *Proceedings. Mathematical, Physical, and Engineering Sciences*, 476(2236), 20190737.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794.

Hasenjager, M. J., Leadbeater, E., & Hoppitt, W. (2021). Detecting and quantifying social transmission using network-based diffusion analysis. *Journal of Animal Ecology*, 90(1), 8–26.

He, P., Klarevas-Irby, J. A., Papageorgiou, D., Christensen, C., Strauss, E. D., & Farine, D. R. (2022). A guide to sampling design for gps-based studies of animal societies. *Methods in Ecology and Evolution*, 1–19. https://doi.org/10.1111/2041-210X.13999

Hinde, R. A. (1976). Interactions, relationships and social structure. *Man*, 11(1), 1–17.

Hobaiter, C., Poisot, T., Zuberbühler, K., Hoppitt, W., & Gruber, T. (2014). Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biology*, 12(9), e1001960.

Hoppitt, W. J., & Farine, D. R. (2018). Association indices for quantifying social relationships: How to deal with missing observations of individuals or groups. *Animal Behaviour*, 136, 227–238.

Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. (2021). Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 5149–5169.

Humle, T. (2011a). The 2003 epidemic of a flu-like respiratory disease at Bossou. In *The chimpanzees of Bossou and Nimba* (pp. 325–333). Springer.

Humle, T. (2011b). Location and ecology. In *The chimpanzees of Bossou and Nimba* (pp. 13–21). Springer.

Jacoby, D. M. P., Fear, L. N., Sims, D. W., & Croft, D. P. (2014). Shark personalities? Repeatability of social network traits in a widely distributed predatory fish. *Behavioral Ecology and Sociobiology*, 68(12), 1995–2003.

James, R., Croft, D. P., & Krause, J. (2009). Potential banana skins in animal social network analysis. *Behavioral Ecology and Sociobiology*, 63(7), 989–997.

Krause, J., James, R., Franks, D. W., & Croft, D. P. (2015). *Animal social networks*. Oxford University Press.

Krause, J., & Ruxton, G. D. (2002). *Living in groups*. Oxford University Press.

Lacreuse, A., Raz, N., Schmidtke, D., Hopkins, W. D., & Herndon, J. G. (2020). Age-related decline in executive function as a hallmark of cognitive ageing in primates: An overview of cognitive and neurobiological studies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 375(1811), 20190618.

Langergraber, K., Mitani, J., & Vigilant, L. (2009). Kinship and social bonds in female chimpanzees (*Pan troglodytes*). *American Journal of Primatology*, 71(10), 840–851.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139.

Machanda, Z. P., & Rosati, A. G. (2020). Shifting sociality during primate ageing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 375(1811), 20190620.

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

Matsuzawa, T. (1994). Field experiments on use of stone tools by chimpanzees in the wild. In F. B. M. de Waal, R. W. Wrangham, W. C. Mcgrew, & P. G. Heltne (Eds.), *Chimpanzee cultures* (pp. 351–370). Harvard University Press.

Matsuzawa, T., Humle, T., & Sugiyama, Y. (2011). *The chimpanzees of Bossou and Nimba*. Springer Science & Business Media.

McCann, R., Bracken, A. M., Christensen, C., Fürtbauer, I., & King, A. J. (2021). The relationship between GPS sampling interval and estimated daily travel distances in chacma baboons (*Papio ursinus*). *International Journal of Primatology*, 42(4), 589–599.

McCarthy, M. S., Després-Einspenner, M.-L., Farine, D. R., Samuni, L., Angedakin, S., Arandjelovic, M., Boesch, C., Dieguez, P., Havercamp, K., Knight, A., Langergraber, K. E., Wittig, R. M., & Kühl, H. S. (2019). Camera traps provide a robust alternative to direct observations for constructing social networks of wild chimpanzees. *Animal Behaviour*, 157, 227–238.

McKinney, W. (2011). pandas: A foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9), 1–9.

Miele, V., Dussert, G., Spataro, B., Chamaillé-Jammes, S., Allainé, D., & Bonenfant, C. (2021). Revisiting animal photo-identification using

deep metric learning and network analysis. *Methods in Ecology and Evolution*, *12*(5), 863–873.

Mitani, J. C. (2009). Male chimpanzees form enduring and equitable social bonds. *Animal Behaviour*, *77*(3), 633–640.

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for gaussian and non-gaussian data: A practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society*, *85*(4), 935–956.

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176.

Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2021). A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, *12*(1), 150–161.

O'Connell, A. F., Nichols, J. D., & Ullas Karanth, K. (2010). *Camera traps in animal ecology: Methods and analyses*. Springer Science & Business Media.

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science Engineering*, *9*(3), 10–20.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems* (Vol. 32). NeurIPS Foundation.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rosati, A. G., Hagberg, L., Enigk, D. K., Otali, E., Emery Thompson, M., Muller, M. N., Wrangham, R. W., & Machanda, Z. P. (2020). Social selectivity in aging wild chimpanzees. *Science*, *370*(6515), 473–476.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113.

Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D., & Carvalho, S. (2019). Chimpanzee face recognition from videos in the wild using deep learning. *Science Advances*, *5*(9), eaaw0736.

Schofield, D. P., Albery, G. F., Firth, J. A., Mielke, A., Hayashi, M., Matsuzawa, T., Biro, D., & Carvalho, S. (2023). Deepprimatenetworksmee.

Sheldon, B. C., Kruuk, L. E., & Alberts, S. C. (2022). The expanding value of long-term studies of individuals in the wild. *Nature Ecology & Evolution*, *6*, 1–3.

Shizuka, D., & Johnson, A. E. (2019). How demographic processes shape animal social networks. *Behavioral Ecology*, *31*(1), 1–11.

Silk, J. B., & Kappeler, P. M. (2017). Sociality in primates. In D. R. In & P. A. Rubenstein (Eds.), *Comparative Social Evolution* (pp. 253–283). Cambridge University Press.

Silk, M. J., Jackson, A. L., Croft, D. P., Colhoun, K., & Bearhop, S. (2015). The consequences of unidentifiable individuals for the analysis of an animal social network. *Animal Behaviour*, *104*, 1–11.

St Clair, J. J. H., Burns, Z. T., Bettaney, E. M., Morrissey, M. B., Otis, B., Ryder, T. B., Fleischer, R. C., James, R., & Rutz, C. (2015). Experimental resource pulses influence social-network dynamics and the potential for information flow in tool-using crows. *Nature Communications*, *6*, 7197.

Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., & Crofoot, M. C. (2015). Shared decision-making drives collective movement in wild baboons. *Science*, *348*(6241), 1358–1361.

Sueur, C., Jacobs, A., Amblard, F., Petit, O., & King, A. J. (2011). How can social network analysis improve the study of primate behavior? *American Journal of Primatology*, *73*(8), 703–719.

Sugiyama, Y. (1984). Population dynamics of wild chimpanzees at Bossou, Guinea, between 1976 and 1983. *Primates*, *25*(4), 391–400.

Sugiyama, Y. (1988). Grooming interactions among adult chimpanzees at Bossou, Guinea, with special reference to social structure. *International Journal of Primatology*, *9*, 393–407.

Sugiyama, Y. (2004). Demographic parameters and life history of chimpanzees at Bossou, Guinea. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *124*(2), 154–165.

Sugiyama, Y., & Fujita, S. (2011). The demography and reproductive parameters of Bossou chimpanzees. In T. Matsuzawa, T. Humle, & Y. Sugiyama (Eds.), *The chimpanzees of Bossou and Nimba* (pp. 23–34). Springer.

Sugiyama, Y., & Koman, J. (1979). Tool-using and -making behavior in wild chimpanzees at Bossou, Guinea. *Primates*, *20*(4), 513–524.

Tkaczynski, P. J., Mielke, A., Samuni, L., Preis, A., Wittig, R. M., & Crockford, C. (2020). Long-term repeatability in social behaviour suggests stable social phenotypes in wild chimpanzees. *Royal Society Open Science*, *7*(8), 200454.

Vallejo Seco, G., Ato García, M., Fernández García, M. P., & Livacic Rojas, P. E. (2013). Multilevel bootstrap analysis with assumptions violated. *Psicothema*, *25*(4), 520–528.

Vidal, M., Wolf, N., Rosenberg, B., Harris, B. P., & Mathis, A. (2021). Perspectives on individual animal identification from biology and computer vision. *Integrative and Comparative Biology*, *61*(3), 900–916.

Wang, W., Neuman, E. J., & Newman, D. A. (2014). Statistical power of the social network autocorrelation model. *Social Networks*, *38*, 88–99.

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature*, *582*(7812), 337–340.

Weinstein, B. G. (2018). A computer vision for animal ecology. *The Journal of Animal Ecology*, *87*(3), 533–545.

Wuerz, Y., & Krüger, O. (2015). Personality over ontogeny in zebra finches: Long-term repeatable traits but unstable behavioural syndromes. *Frontiers in Zoology*, *12*(Suppl 1), S9.

Yamakoshi, G. (1998). Dietary responses to fruit scarcity of wild chimpanzees at Bossou, Guinea: Possible implications for ecological importance of tool use. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *106*(3), 283–295.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Kernel density plot for distribution of individual-level strength values.

Figure S2. Output from normality checks using the 'check_model' function in the PERFORMANCE package in R (Lüdecke et al., 2021) for the individual strength variable.

Figure S3. Kernel density plot for distribution of individual-level Mean Group Size values.

Figure S4. Output from normality checks using the 'check_model' function in the PERFORMANCE package in R (Lüdecke et al., 2021) for the Mean Group Size variable.

Figure S5. Kernel density plot for distribution of individual-level 'Time Alone' values.

Figure S6. Output from normality checks using the 'check_model' function in the PERFORMANCE package in R (Lüdecke et al., 2021) for the individual-level Time Alone variable. Normality checks of Time Alone reveal that the assumption of normality has been violated

with a slightly longer tail than would be expected from a normally distributed variable and beta distribution, indicating our model is systematically underestimating the upper tail end of our distribution (see distribution of predicted values below). However, this may be due to the nature of our dataset which has some outliers/inflated proportion scores, above that which would be expected for a chimpanzee (i.e. above 10% of their time spent alone). Furthermore, it has been demonstrated that violation of normality assumptions does not significantly affect regression estimates for mixed models (Vallejo Seco et al., 2013). Although the reference line for homogeneity of variance is not perfectly horizontal, overall there is an equal spread around this line indicating that the amount of heteroskedasticity is unproblematic.

Figure S7. Bossou Age distribution (scaled to be within 0 and 1) by Sex. Across years, there is a small proportion of middle-aged individuals, with a large core of elderly females.

Figure S8. Model outputs from data generated using intermediate 120s sampling/aggregation rate, visualising the effect of Age on each outcome variable (left column). Corresponding forest plots (right column) showing the effect of each fixed effect (SexM—effect of Males compared to Females), Age (fitted as a quadratic term) and population size (as a controlled fixed effect). Point estimates denote the effect size between each variable on outcome variable, error bars the 95% confidence intervals and asterisks statistical significance (with estimated confidence intervals not overlapping with zero). Model parameters are detailed in Tables 1 and 2. Note that visualised model fit for Age for Strength and Mean Group Size (left) are non-significant. To assess whether variables of interest (Age and Sex) significantly impact our model's predictive ability, check significance of likelihood ratio tests which are detailed in Table 2.

Figure S9. A zoom shot of two old females at the outdoor lab in 2012 eating palm fruit (left: Velu, right: Yo) (photo credit: Tetsuro Matsuawa).

Table S1. Information on videos for each year in the dataset. 'Total observation time' refers to the hours of video data sampled for each year and 'mean duration' refers to the average length of the video clips.