Observation methods in animal behaviour: a simulation study of performance

Alexander Mielke[1], Camille Testard[2], Alba Motes-Rodrigo[3], Lauren J. N. Brent[4], Delphine De Moor[4,5]

[1] Centre for Brain and Behaviour, School of Biological and Behavioural Sciences, Queen Mary University of London, UK

[2] Harvard Society of Fellows, Department of Molecular and Cellular Biology, Harvard, Cambridge, USA

[3] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[4] School of Psychology, Centre for Research in Animal Behaviour, University of Exeter, Exeter, UK

[5] Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Corresponding author: Alexander Mielke, G.E. Fogg Building, Mile End Rd, E14NS, London, UK; a.mielke@qmul.ac.uk

Keywords

Observation methods, focal follows, group scan, continuous sampling, time sampling, behavioural sampling, study design, data collection

Abstract

Most behavioural studies rely on systematic samples of behaviour, as observing and recording all behavioural events that occur is rarely feasible. Choosing an observation method involves several key decisions, including which individuals to observe, how to sample their behaviour, and how to distribute sampling effort over time. These decisions influence how closely behavioural estimates reflect the true occurrence of behaviours and how comparable estimates are across studies using different methods. Here, we used a simulation approach to evaluate the performance of different observation methods in terms of how accurately and precisely they represent true behavioural occurrences across varying contexts. We simulated behaviours differing in duration, frequency, and observability, in animals living in groups of different sizes and terrains with varying visibility. We then tested how the two most common observation methods—focal follows and group scans—captured these behaviours across different study durations, scan intervals, and focal lengths. We found that focal follows generated more accurate and precise behavioural estimates for short, rare behaviours, while group scans performed better for longer, more common behaviours. Group scans also performed better in larger group sizes and shorter study durations, as long as a large proportion of individuals was visible. We provide researchers with an interactive tool, the SIMBO app, to explore which observation method might be best suited to the specific properties of their system and research question. Overall, our study and app offer quantitative guidance on the performance of focal follows and group scans across contexts and highlights potential pitfalls for comparative research using data collected with different methods.

Introduction

Behaviour is a key interface between animals and their environment, shaping their inclusive fitness and driving evolutionary change (Kappeler et al., 2013; Leimar et al., 2022). Much of our understanding of animal behaviour comes from direct observation. However, because it is rarely feasible to record all behavioural events that occur, researchers typically rely on systematic samples of behaviours collected using established observation methods (Altmann, 1974; Bateson & Martin, 2021). Designing behavioural observation studies therefore involves a series of critical decisions: which individuals to observe, how to sample their behaviour, and how to distribute sampling effort over time (Fragaszy et al., 1992). These choices are not merely technical—they are central to optimizing the statistical power, reliability, validity, and generalizability of the resulting behavioural estimates (Fragaszy et al., 1992; Mielke et al., 2021). Understanding how sampling choices impact behavioural estimates is also crucial for interpreting differences across studies and for conducting reliable comparative analyses (De Moor et al., 2024; Webster & Rutz, 2020).

Simply recording all behaviours that observers can see (a method known as '*ad libitum* sampling') can introduce biases, as more conspicuous behaviours and individuals tend to be recorded more frequently than less noticeable ones (Aldrich-Blake, 1970; Altmann, 1974). To address these biases, several systematic sampling methods have been developed (Altmann, 1974; Whitehead, 2008). These methods are primarily defined by two factors: which individuals are observed ('sampling rules') and how the behaviour is recorded ('recording rules'; Bateson & Martin, 2021). The most common sampling rules focus on observing either a single individual at a time (focal sampling) or a group of individuals (group sampling). For recording behaviour, two main approaches are widely used (Fig. 1): continuous sampling and time sampling. Continuous sampling captures all occurrences of a behaviour over a fixed period, either as counts or durations of behaviours. Time sampling, on the other hand, records whether one or several behaviours are occurring or not at regular intervals, yielding binary data.

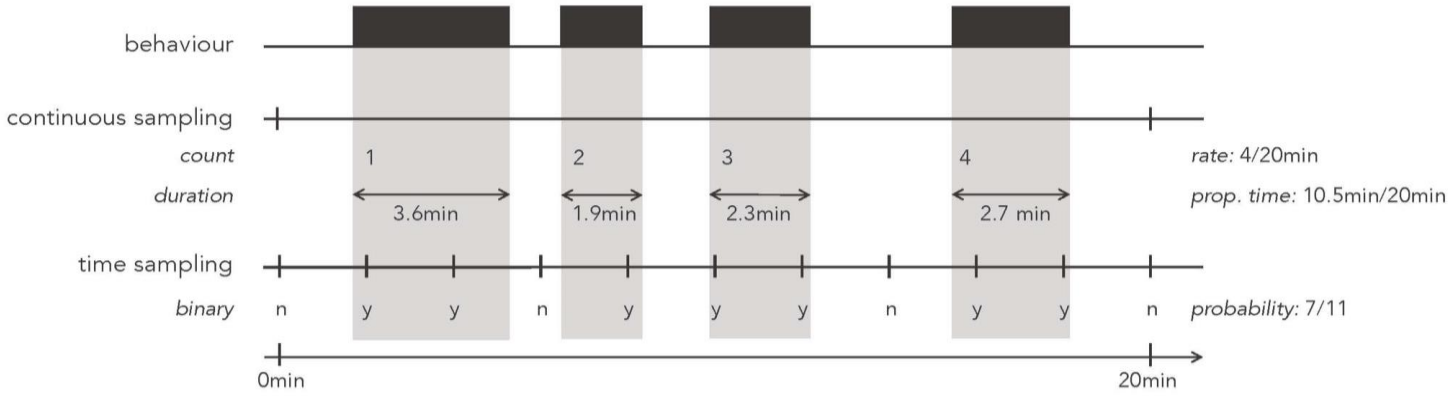

Figure 1. *Schematic visualisation of two primary recording rules. "Continuous sampling" records every occurrence of a behaviour within a predefined observation time, either as a count of events or the total duration of the behaviour. From these data, behaviour rates (count/observation time) or proportions of time (duration/observation time) can be calculated. In contrast, time sampling records whether a behaviour is occurring at regular intervals, resulting in binary data, from which a probability (number of samples with behaviour/total number of samples) can be calculated. For the same underlying behaviour (top row in black), continuous sampling of event count gives an estimated rate of 0.2 events per minute (4/20), continuous sampling of durations results in an estimated proportion of time of 0.52 (10.5/20) and time sampling yields an estimated probability of 0.63 (7/11), highlighting that the same behaviour gives rise to different behavioral estimates based on the recording rule that is used.*

Focal sampling is typically paired with continuous recording, allowing researchers to record all behaviours of the focal individual, including their timing and sequence. In contrast, group sampling is often combined with time sampling, as continuously recording all occurrences of a behaviour in a group is usually not feasible. Consequently, focal continuous sampling (from here on referred to as '**focal follows**') and group time sampling (from here on referred to as '**group scans**') are two of the most widely used observation methods in animal behaviour research (Brereton et al., 2022; Webber & Vander Wal, 2019; Whitehead, 2008).

Each of these observation methods has its own strengths and limitations. Focal follows provide rich, detailed data about the focal individual but overlook all behaviours not involving that individual. On the other hand, group scans provide a broader overview of behaviour across group members at various times of the day but miss sequences of behaviours and behaviours that occur between sampling intervals. The performance

of each observation method might also depend on several factors related to the study system, the behaviour of interest, and the observational setup. For instance, group size determines the amount of data that can be collected per individual within a given timeframe and visibility of the terrain impacts how well the entire group can be observed at once: large groups in open savannahs, for example, present very different observational challenges than small groups in dense rainforests. The duration, frequency and visibility of the behaviour of interest can also impact how likely it is that a behaviour is recorded. Shorter and less visible behaviours such as facial expressions are more easily missed than conspicuous fights, and rare behaviours are easier to miss compared to common ones. Finally, aspects of the observational setup, such as the duration of focal follows, the intervals between scan samples, and the time spent observing each individual during a scan, can also affect performance. Each of these variables can influence how well the recorded data represent the true occurrence of behaviour of interest, and thus which observation method is best suited for a given study system or research question.

Understanding how focal follows and group scans perform across contexts is also crucial for comparative studies, where data collected using various methods are often combined (De Moor et al., 2024; Nunn, 2011; Rubenstein & Abbot, 2017). One major challenge in comparative animal behaviour research is determining whether observed differences in behaviour are due to true biological variation or simply result from differences in observation method (Ihle et al., 2017; Lukas & Clutton-Brock, 2017; O'Dea et al., 2021; Webster & Rutz, 2020). Insights into how similarly or differently observation methods perform can help researchers assess when comparisons are valid and when methodological differences may bias their findings.

Studies providing guidance on observation methods and their respective strengths and weaknesses have been highly influential (Altmann, 1974; Bateson & Martin, 2021; Lehner, 1998; Whitehead, 2008), and several studies have estimated their relative performance (e.g. Amato et al., 2013; Brereton et al., 2022; Canteloup et al., 2020; Castles et al., 2014; Davis et al., 2018; Fragaszy et al., 1992; Gilby et al., 2010; Hämäläinen et al., 2016; Hepworth & Hamilton, 2001; Karniski et al., 2014; Pullin et al., 2017; Rose, 2000). Yet, no study to date has evaluated the extent to which performance of observation methods is influenced by system-, behaviour-, and

observational setup-specific factors. A main reason for this is that the true occurrences of a behaviour in real-world systems is rarely known, making it difficult to benchmark behavioural estimates against a ground truth. Simulation studies are a good way to address this issue, because they allow us to set group-specific, behaviour-specific, and observation-specific parameters with full knowledge of the true occurrences of behaviour. While the resulting data and comparisons are much simpler than a true biological system, they nevertheless can act as guidance for researchers designing a behavioural study (Fogarty et al., 2022; Williams et al., 2024).

Here, we conducted a simulation study to test the performance of focal follows and group scans. We do this by quantifying how accurately and precisely the data recorded using these observation methods estimate the true occurrence of simulated behaviours, as well as the comparability of the estimates they produce. We also evaluate how factors related to the system, behaviour and observational setup affect the performance of both methods. Our goal is to offer rigorous, quantitative guidance for selecting the observation method for a given study. Additionally, we aim to help researchers assess the extent to which differences in observation methods might drive observed behavioural differences when comparing data.

Methods

Simulations

To create a ground truth of occurrences of behaviour against which to quantify the performance of observation methods, we simulated behaviours occurring in virtual groups of animals. We varied the following parameters to estimate their impact on the performance of focal follows and group scans: *group size*, *terrain visibility* (are all individuals visible at all times?), *behaviour visibility* (is it likely that an observer sees the behaviour when it happens?), *behaviour frequency*, *behaviour duration*, *study duration*, *focal duration*, *focal break time* (simulating the time needed to find a new subject to follow), *scan interval time* and *scan time per subject* (simulating that it might be harder to note down some behaviours or find each individual in the sampling range; see Table 1 for definitions and value spaces). We assumed seven hours of observations on any given day. We randomly selected values from each parameter and combined them, generating a total of ~229,000 simulations.

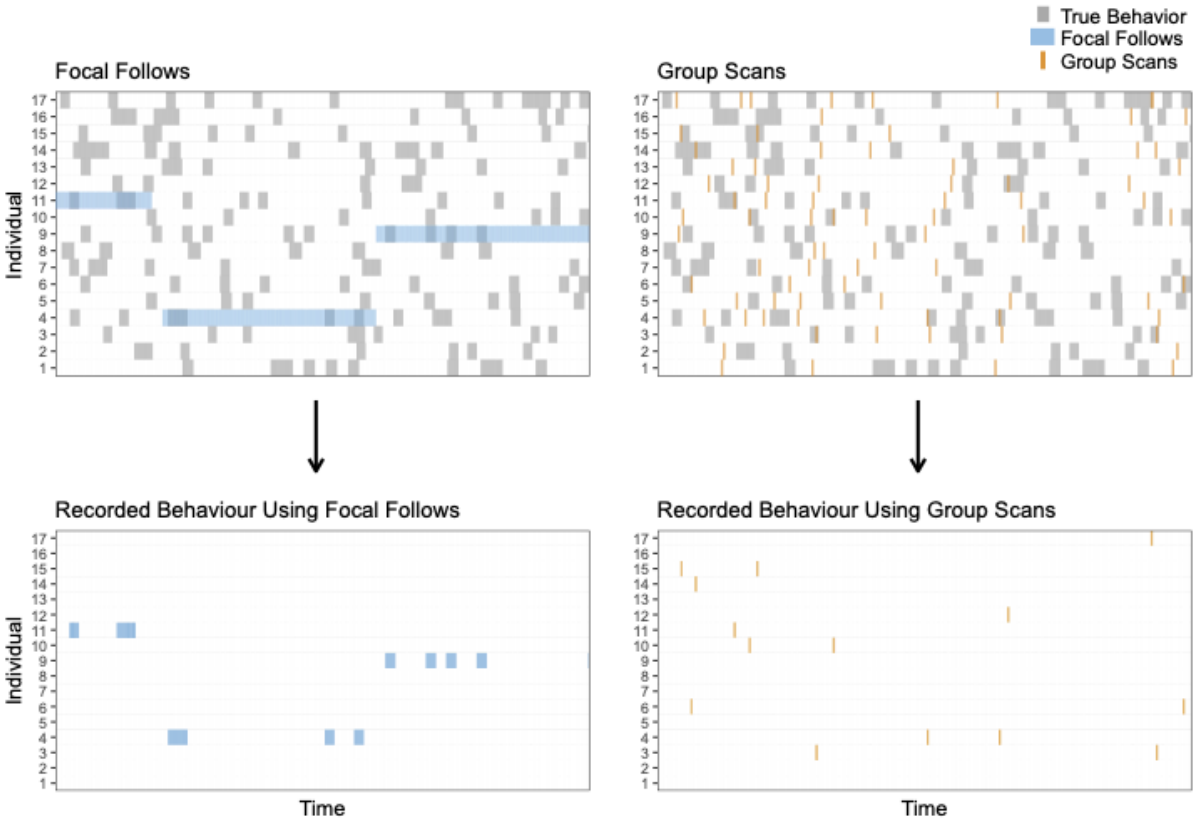

Figure 2. *Visual representation of the simulated behaviour and observation methods. The top two panels show the same true occurrences of behaviour (in grey) for each individual across time. The left panel represents observation using focal follows (in blue), where one individual is observed for a fixed period of time. Below it, the bottom-left panel shows the behaviour that was recorded during these focal follows. On the right side, the top-right panel represents observation using group scans (in yellow), where all individuals are observed near-instantaneously at regular time intervals. The bottom-right panel shows the recorded behaviour from these group scans. This figure illustrates how the same underlying behaviour (grey) can be sampled using two different observation methods, highlighting the differences in the data recorded using focal follows versus group scans.*

In each simulation, we first created a set of seconds for each seven-hour day of the entire *study duration* for a fixed number of individuals (*group size*). Then on each day, we randomly allocated the number of behavioural events (*behaviour frequency*) across individuals. This approach assumes that individuals engage in the behaviour randomly throughout the day and across the study period, and that behavioural occurrences are independent from each other. Individuals had at least one behavioural occurrence per day. The duration of behaviours was fixed to one value (*behaviour duration*). At the

end of each simulation we obtained the true proportion of time each individual spent engaged in a behaviour, which we refer to as the **true value** from here on.

Next, we simulated the observation methods. Focal follows in our simulations lasted for a set number of minutes (*focal duration*), with a set *focal break time* between the end of a given focal follow and the start of the next. For each focal follow, the duration of the behaviour was recorded as the number of seconds in which the individual was engaged in the behaviour. The focal individual was randomly selected from the pool of individuals, with no rule against picking the same individual twice in a row (behaviours are spread out randomly in time, so the order at which individuals are selected to follow does not affect results). For group scans, the number of scans in a day was determined in each simulation by the *scan interval time* parameter, which sets the number of minutes between the end of one scan and the start of the next. During a scan, individuals were observed sequentially and near-instantaneously, recording whether they were engaged in the behaviour or not. The order at which the visible individuals were scanned was randomly assigned, with a fixed number of seconds spent on each individual (*scan time per subject)*. An individual was considered to be engaged in the behaviour in a given scan if the seconds in which it was scanned contained the behaviour.

In each simulation, we also set two parameters that determine the likelihood of missing a behaviour. The first parameter, t*errain visibility,* represents the fact that individuals in some terrains are harder to keep in view than in others. In focal follows, it is implemented as the proportion of the follow during which the subject is out of view (and thus unobservable). In group scans, it is the proportion of individuals not observed in a given scan. For example, if terrain visibility is set to 0.5, then the subject is effectively “out of view” for half of the focal follow, and half the group members are unseen in each scan. Importantly, this out-of-view period does not count toward actual observation time for focal follows, and group scans are ended after all visible individuals have been scanned, which can result in more scans per day as scans are shorter.

The second visibility parameter, *behaviour visibility,* represents how likely an observer is to observe the behaviour in question, even when an individual is in view. For focal

follows, it is implemented as the proportion of time the focal subject is observed but no behaviour is recorded. For group scans, it is the proportion of individuals that are seen and scanned but have no recorded behaviour. For example, if behaviour visibility is set to 0.5, then no behaviours are recorded for half of the focal follow or half of the group members in each scan, even if behaviours actually occur. Unlike *terrain visibility*, these "missed" intervals count toward the total observation time for focal follows, and scans last as long as it takes to observe each visible individual—reflecting that the observer is watching but that the behaviour is not seen.

Table 1. *Summary of the simulation parameters, including their definitions and the range of values used. Value ranges were determined to represent realistic values for each parameter that researchers might encounter in behavioural studies. Values used in the SIMBO Shiny app (see below) are highlighted in bold.*

Category	**Parameter**	**Definition & implementation**	**Value Range**
Study system	Group size	Number of individuals in the group for which behaviour is recorded..	10, 15, **20**, 25, 30, 35, 40, 45, **50**, 55, 60, 65, 70, 75, 80, 85, 90, 95, **100**
	Terrain visibility	Proportion of individuals visible at any time. For focal follows, this determines the proportion of time during a focal period that the subject is visible for behaviour recording. For group scans, it determines the proportion of all individuals that are visible for behaviour recording.	0.1, **0.2**, 0.3, 0.4, **0.5**, 0.6, 0.7, **0.8**, 0.9, 1
Behaviour	Behaviour frequency	Average number of occurrences of the behaviour per individual per day, with some variation between individuals (SD = 2).	**1**, 2, 3, 4, 5, 6, **7**, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, **20**, 25, 30, 35, 40, 45, **50**
	Behaviour duration	Duration of all occurrences of the behaviour in seconds.	1, 2, **3**, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, **30**, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 110, **120**, 180, 240, 300, 360, 420, 480, 540, **600**

	Behaviour visibility	Proportion of the behavioural occurrences that are visible. For focal follows, this determines the proportion of time during a focal period that no behaviours are recorded, even if they occur. For group scans, it determines the proportion of all individuals for which no behaviour is recorded, even if it occurs.	0.1, **0.2**, 0.3, 0.4, **0.5**, 0.6, 0.7, **0.8**, 0.9, 1
Observational setup	Study duration	Number of observation days, with each day being seven hours long.	**30**, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, **90**, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, **180**
	Focal duration	Length of a focal follow (in minutes).	5, 10, **15**, 20, 25, 30, 40, 50, **60**
	Focal break time	Length of time between the end of a focal follow and the start of the next one (in minutes).	1, 2, 3, 4, **5**
	Scan interval time	Length of time between the end of a group scan and the start of the next one (in minutes).	**1**, **5**, 10, **15**, 20, 25, 30, 40, 50, 60
	Scan time per subject	Time needed to observe each individual during group scans (in seconds).	**1**, 3, **5**, 7, 9, 11

At the end of each simulation, we estimated the proportion of time an individual spent engaged in a behaviour as follows: For focal follows, we calculated the proportion of time an individual spent engaged in a behaviour as the number of seconds an individual was observed engaging in the behaviour divided by the total number of seconds the individual was observed. For group scans, we calculated the individual probability to engage in the behaviour as the number of scans where an individual was observed to engage in the behaviour divided by the total number of times they were observed in a scan. The proportion of time spent in a particular behaviour is equivalent to the probability of being in that state at a given point in time, allowing direct comparisons of the performance of both observation methods. From here on we refer to these estimated proportions/probabilities as the **behavioural estimates**.

In total, there were 25,304,853,600 possible combinations of all simulation parameters. Due to the computational expenses of simulating each of those, we ran 65,620 simulations randomly selecting one value for each parameter. For each set of parameters, we produced 10 iterations of the simulations, to quantify how similar the estimates were when using the same observation method, given that random factors vary between runs.

How do focal follows and group scans perform?

To evaluate the performance of focal follows and group scans, we computed two main measures: accuracy and precision. We calculated two additional measures, bias and correlation to true value, which we did not include in all investigations, as they were mainly intended to provide additional context on the accuracy and precision of the behavioral estimates.

1. **Accuracy:** how closely do behavioural estimates match the true value?

We measured accuracy as the standardized root mean squared error (RMSE, Formula 1) between each individual's behavioural estimate and its true value across iterations of the same simulation. Because larger true values naturally produce larger RMSE, we standardized RMSE by dividing it by the true value. As such, the standardized RMSE represents the average error as a percentage of the true value: an RMSE of 0 indicates a perfect match between the behavioral estimate and the true value, and values greater than 100 mean that the error is larger than the true value itself. The RMSE is not sensitive to direction, i.e., we do not know whether the behavioural estimate over- or underestimates the true value. However, because the observed values are bound by 0 for each individual, RMSE values above 100 can only arise if the error was an overestimation.

$$Standardised\ RMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}}{y}$$

***Formula 1**: Standardised Root Mean Squared Error as a measure of accuracy. ŷ are the behavioural estimates, y are the true values, n are the number of individual estimates.*

2. **Precision:** how consistent are behavioural estimates of the same true value?

We measured precision as the coefficient of variation (CV, Formula 2) of each individual's behavioural estimate across the ten iterations of each simulation with a given set of parameters. As such, precision indicates how similar a new estimate would be if the same behaviour was observed again under identical conditions. A lower CV indicates more consistent (i.e., more precise) estimates, meaning a new estimate under identical conditions would likely be similar. In some cases the CV was 0 because all iterations estimated a value of 0, reflecting a uniformly incorrect outcome (all iterations missed the true occurrence of the behaviour) rather than true precision. We replaced these 0 values with missing data to maintain interpretability, ensuring that smaller CV values actually represent higher precision.

$$CV = \frac{\sigma}{\mu}$$

***Formula 2**: Coefficient of variation as a measure of precision. σ is the standard deviation of an individual's behavioural estimates across the ten iterations of the simulation, and μ is the mean of those estimates.*

3. **Bias:** do behavioural estimates over- or underestimate the true value?

We measured bias as the standardized mean error between each individual's behavioural estimate and its true value across iterations of the same simulation (Formula 3). Because larger true values naturally produce larger differences, we standardized bias by dividing it by the true value. A bias of 0 indicates that, on average, the estimate matches the true value, while positive values indicate overestimation and negative values indicate underestimation. A method can generate estimates with minimal bias that are still inaccurate (i.e., show large random variation around the true value). Therefore, combining accuracy and bias provides a more complete view of how closely behavioural estimates align with the true values.

$$Standardised\ Mean\ Error = \frac{1}{n}\sum\frac{(y_i - \bar{y}_i)}{y_i}$$

Formula 3*: Standardized mean error as a measure of bias. ŷ are the behavioural estimates, y are the true values, n are the number of individual estimates.*

4. **Correlation to true value**: are behavioural estimates ranked similarly to true values?

For many questions in animal behaviour research, the primary goal is to assess how individuals differ from each other, rather than to obtain precise and accurate measurements of each individual's behaviour. For example, researchers might be interested in identifying which animals display the highest levels of aggression or spend the most time feeding, rather than focusing on the exact amount of time each individual spends on those behaviours. To test how focal follows and group scans perform in correctly representing the ranking of individuals relative to others, we calculated Spearman rank correlations between the behavioural estimates and true values. We calculated correlations for each iteration of the simulation and then averaged them across iterations.

Do focal follows and group scans produce comparable results?

To estimate how comparable data collected using focal follows and group scans were, we calculated the Spearman rank correlation between the behavioural estimates produced by each method. We calculated correlations for each iteration of the simulation and then averaged them across iterations.

How do system, behaviour, and observation parameters impact method performance?

To estimate how factors related to the study system, the behaviour of interest, and the observational setup influence the performance of different observation methods, we fitted models with accuracy and precision as response variables — separately for focal follows and group scans. As predictors, we included the parameters detailed in Table 1, z-standardising them before inclusion to ensure improved model convergence and interpretability (Schielzeth, 2010). Because accuracy and precision values were

heavily right-skewed, we modelled them as log-normal using the lme4 package (Bates et al., 2015). For the focal follow models, we excluded scan-specific predictors (*scan time per subject* and *scan interval time*), and for group scan models, we excluded focal-specific predictors (*focal duration* and *focal break time*). We included simulation ID as a random effect.

These models give an indication of the magnitude and direction of each parameter's effect on method performance. However, in reality, these parameters interact in complex ways that are difficult to fully capture through modelling alone. To address this, we present a decision tree, a set of illustrative case studies and an interactive Shiny app that allows researchers to explore the effects of different parameter combinations.

Which parameters drive differences between group scans and focal follows?

To decide whether to choose focal follows or group scans, researchers may wish to know under which sets of parameters one observation method performs better than the other. Here, we investigated this using a classification approach: for each simulation, we calculated the median accuracy values across all simulated individuals, in line with the idea that researchers would want to find the approach that represents all group members most accurately. We classed each simulation as 'focal better' if focal follows showed lower standardised RMSE; 'scan better' if group scans showed lower standardised RMSE; and 'same' if their standardised RMSE values were within 1% error of each other. We used random forests as a classification algorithm, implemented in the 'rpart' package (Therneau et al., 2025) with default hyperparameter settings, and plotted the pruned decision tree of the forest. We present the decision tree based on accuracy, because the pattern did not differ for precision, and because accuracy is usually the measure that researchers want to optimise the most. We interpret any splits that separate the performance of the two approaches from each other.

Case studies

We present four case studies and use our simulations to investigate the impact of choosing one observation method over the other by answering common questions researchers may face before starting their study.

1. ***What observation method should I use to record my behaviour of interest?***

We tested the performance of focal follows and group scans in capturing three types of behaviours:

- A long, common, and highly visible behaviour like grooming or traveling (*behaviour frequency* = 10 occurrences per day per individual, *behaviour duration* = 60 seconds, *behaviour visibility* = 0.9).
- A short, less common, and highly visible behaviour like aggression or drinking (*behaviour frequency* = 5 occurrences, *behaviour duration* = 3 seconds, *behaviour visibility* = 0.9 behaviour visibility).
- A short, less common, less visible behaviour like a threat posture or scratching (*behaviour frequency* = 5 occurrences, *behaviour duration* = 3 seconds, *behaviour visibility* = 0.3).

We fixed the remaining parameters to *study duration* = 200 observation days, *group size* = 90 individuals, *terrain visibility* = 0.5, *focal duration* = 15 minutes, *focal break time* = 1 minute, *scan interval time* = 5 minutes and s*can time per subject* = 3 seconds.

2. ***Is the relative performance of group scans versus focal follows influenced by group size?***

We tested the performance of group scans relative to focal follows for groups of different sizes (15, 50, or 90 individuals). We parameterized the behaviour as a long, common and highly visible behaviour like grooming or traveling (*behaviour frequency* = 10 occurrences per day per individual, *behaviour duration* = 60 seconds, *behaviour visibility* = 0.9). We fixed the remaining parameters to *study duration* = 200 observation days, *terrain visibility* = 0.5, *focal duration* = 15 minutes, *focal break time* = 1 minute, *scan interval time* = 5 minutes and s*can time per subject* = 3 seconds.

3. ***How long does my study need to be?***

Assuming that individuals have a stable tendency to engage in a given behavior over time, longer studies should produce behavioural estimates that more closely reflect the true value. We tested whether different *study durations* (30 days, 90 days, 180 days, 730 days) were associated with different performance for focal follows and group scans respectively. We tested this for two types of behaviours:

- A long, common, and highly visible behaviour like grooming or traveling (*behaviour frequency* = 10 occurrences per day per individual, *behaviour duration* = 60 seconds, *behaviour visibility* = 0.9).
- A short, less common, and highly visible behaviour like aggression or drinking (*behaviour frequency* = 5 occurrences, *behaviour duration* = 3 seconds, *behaviour visibility* = 0.9 behaviour visibility).

We fixed the remaining parameters to *group size* = 40 individuals, *terrain visibility* = 0.5, *focal duration* = 15 minutes, *focal break time* = 1 minutes, *scan interval time* = 5 minutes and s*can time per subject* = 3 seconds.

SIMBO Shiny App

To allow researchers to fully explore the parameter space of our simulations, we developed a user-friendly Shiny app, 'SIMBO' (Simulator of Methods for Behavioural Observation: anonymized link). For the SIMBO app, we generated a further ~31,000 simulations of all possible combinations of a predefined subset of parameter values within their expected common ranges. This approach allowed us to fully represent every possible combination within that subset, making the app interactive and comprehensive within a manageable scope. In contrast, the simulations used for our statistical analyses were based on random draws across the entire parameter space to ensure broad coverage, but did not include all possible combinations. The SIMBO app enables users to investigate how various parameters related to the study system (*group size*, *terrain visibility*), behaviour of interest (*behaviour visibility, behaviour frequency*, *behaviour duration*) and observation method (*study duration, focal duration*, *scan interval time*, *scan time per subject*) influence the accuracy and precision of behavioural estimates from focal follows and group scans. As such, the SIMBO app enables researchers to assess what observation method might be best suited to the specific properties of their system and research question.

Results

How do focal follows and group scans perform?

Considering behavioural estimates drawn across all investigated parameters, focal follows estimated true values with slightly lower error (standardised RMSE = 54.67) than group scans (standardised RMSE = 63.89). For both observation methods, the distribution of accuracy peaked at a standardised RMSE of 100 (i.e. an error as large as the true value), indicating that individual values were often missed, i.e. estimated as zero even though the behaviour did take place (Fig. 3A). The median precision was similar for focal follows (CV 35.40) and for group scans (CV = 34.17; Fig. 3B), as was the median correlation between the behavioural estimates and true values (correlation = 0.41 for focal follows and 0.39 for group scans, respectively; Fig. 3D).

Thus, on average across all simulations, focal follows generated slightly more accurate behavioural estimates than group scans (Fig. 3A). The underlying cause of this difference is explained by the negative bias values (Fig. 3C): behavioural estimates from focal follows rarely overestimated true values, so error values cluster between 0 and the true value, which limits the maximum error that can be observed. This is also visible as the strong cutoff of the distribution in Fig. 3A, with almost no standardised RMSE > 100 for focal follows. The distributions of precision and accuracy for group scans are broader, meaning that scans can generate more precise and accurate behavioural estimates when they work well, but also have the potential to strongly overestimate true values. Thus, the respective benefits of focal follows and group scans are dependent on the data collection context.

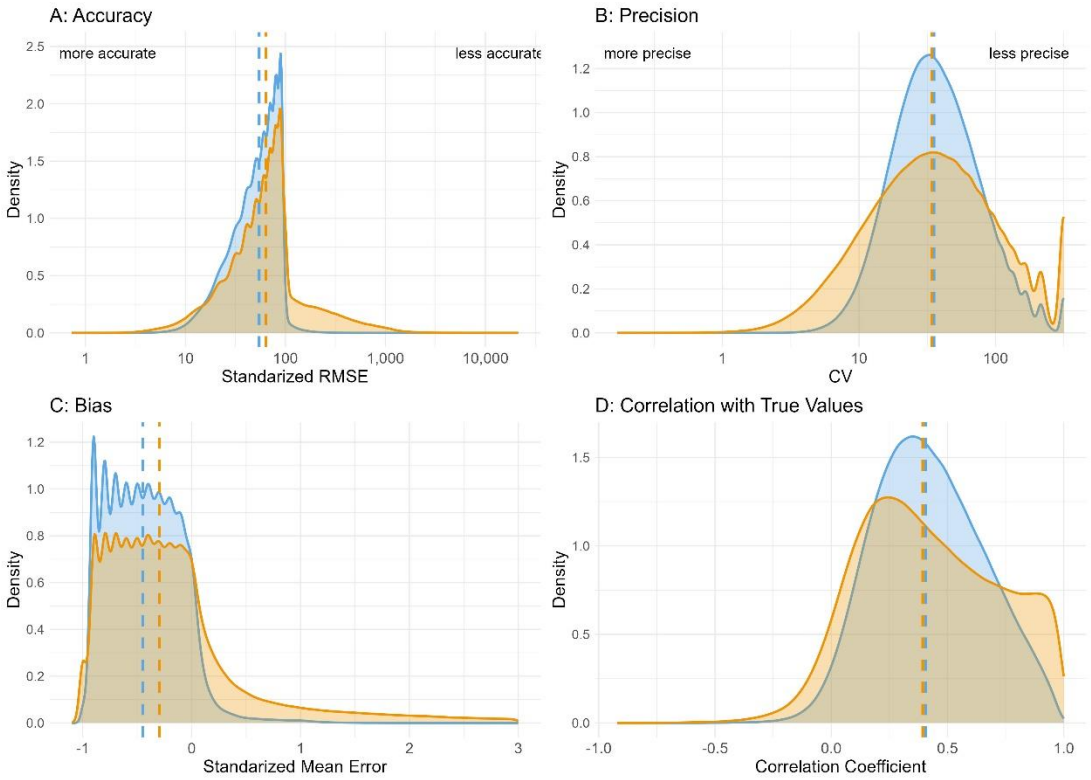

***Figure 3**: Distributions of A: standardized RMSE values as a measure of accuracy; B: CV values as a measure of precision across multiple iterations of the same set of simulation parameters; C: standardized mean error values as a measure of bias and D: correlation values between the true value and behavioural estimates. Dashed lines indicate medians. Distributions in blue represent measures for behavioural estimates from focal follows, distributions in yellow represent measures for behavioural estimates from group scans. The X-axis for A and B is log-scaled to accommodate large values. In plots A, B and C, measures closer to 0 indicate higher accuracy, precision, and bias, respectively, and therefore a better representation of the true behavioural values. In plot D, measures closer to 1 indicate a higher correlation with the true value.*

Do focal follows and group scans produce comparable results?

When comparing behavioural estimates from focal follows and group scans of the same true behaviour, we found that the correlation between the two was generally positive but low, with a median correlation coefficient of 0.18 (Fig. 4). Therefore, for all but a small subset of comparisons, the two methods produced different behavioural estimates indicating that in most cases the results obtained from both observation methods cannot be directly compared.

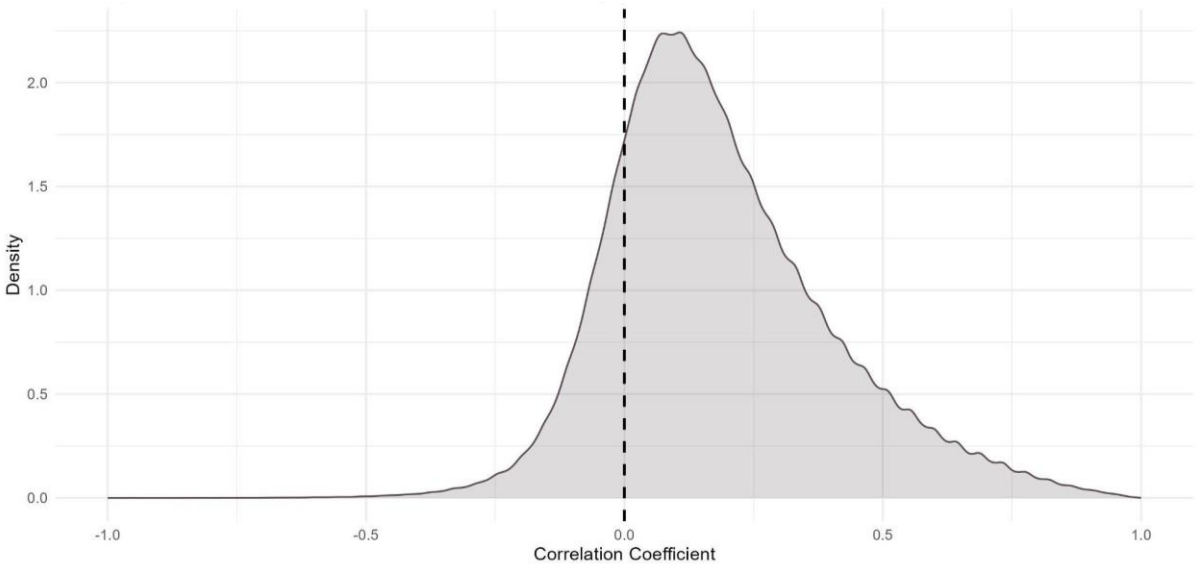

Figure 4: *Distribution of correlation values between the behavioural estimates of focal follows and group scans from the same true value. Values closer to 1 indicate that the two observation methods produced behavioural estimates similar to one another.*

How do system, behaviour, and observation parameters impact method performance?

Parameters of the study systems impacted focal follows and group scans differently (Fig. 5). For focal follows, the accuracy and precision of behavioural estimates were lower for larger *group sizes*. This is likely because the total available observation time needs to be divided among more individuals, reducing the amount of time spent observing each one, and therefore more behaviours per individual were missed. In contrast, *group size* did not directly affect the accuracy or precision of group scans. Instead, the accuracy of group scan behavioural estimates was more strongly influenced by the *terrain visibility*, with a higher proportion of individuals visible during each scan leading to more accurate estimates. *Terrain visibility* had minimal impact on focal follow performance, which is likely to be the case because periods during which focal individuals were out of sight did not count toward observation time.

Behavioural parameters strongly influenced the accuracy and precision of behavioural estimates from both focal follows and group scans (Fig. 5). Estimates of behaviours with a higher *behaviour frequency* and *behaviour duration* were more accurate and especially more precise for both focal follows and group scans, as they were less likely to be missed. *Behaviour duration* impacted the performance of group scans more

strongly, because behaviours of short durations were more easily missed if they occurred in between scans, while behaviours occurring during focal follows were recorded regardless of their duration. Higher *behaviour visibility* also improved accuracy and precision of behavioural estimates, particularly for focal follows. This is likely because a low-visibility behaviour might be entirely missed during a focal follow, whereas during group scans the behaviour might be recorded in subsequent scans.

Finally, most observational setup parameters had relatively limited effects on the performance of focal follows and group scans. Longer *study durations* slightly improved accuracy and precision for both methods, probably because with longer study durations, more behaviours are recorded and variability is averaged out, leading to more reliable estimates. For focal follows, *focal duration* and *focal break time* had negligible effects on accuracy and precision. For group scans, shorter *scan interval times* (i.e. more frequent scans) improved accuracy and especially precision, likely due to an increased likelihood of detecting short-duration behaviours. Longer *scan time per subject* slightly decreased accuracy, likely because this decreased the total number of scans that could be done per day.

Note that the effects described here represent the influence of each parameter individually, holding all other parameters constant at their mean values. However, in practice, these parameters often co-vary, leading to more complex interactions.

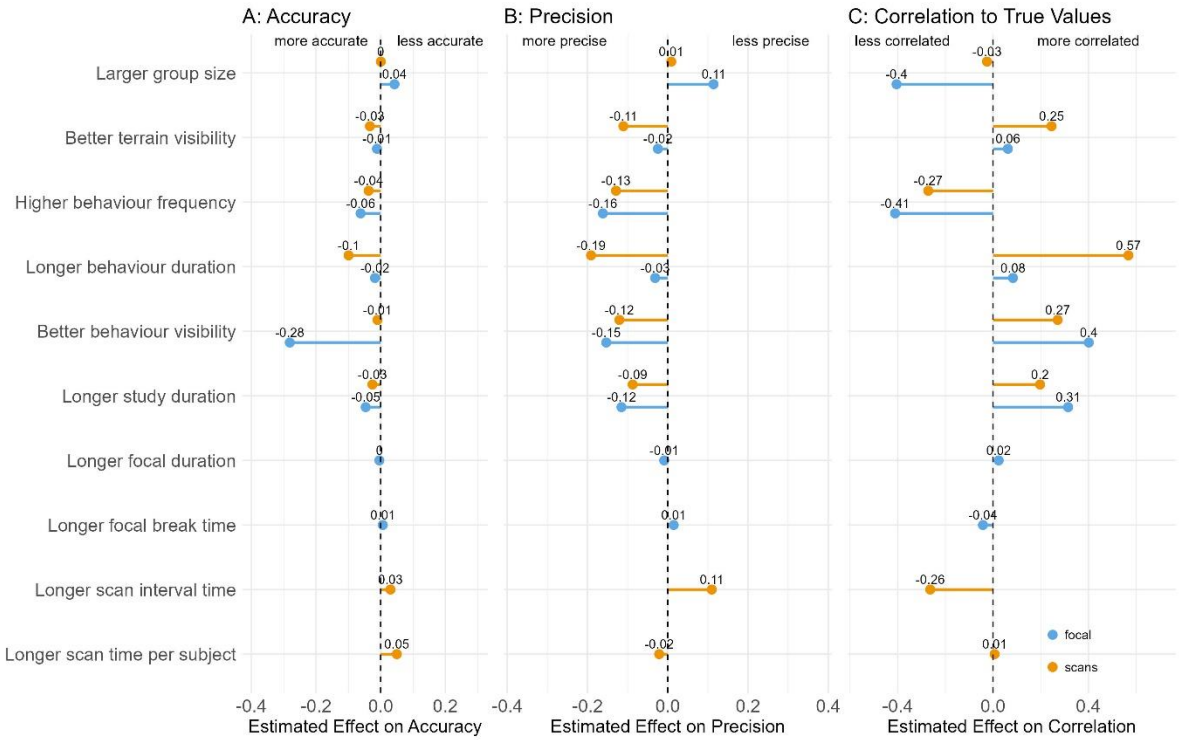

Figure 5*: The estimated effect of system, behaviour and observation parameters on A: the accuracy of behavioural estimates, B: the precision of behavioural estimates and C: correlation values between the true value and behavioural estimates. Model estimates are in blue for focal follows and in yellow for group scans. Lower values indicate higher accuracy (lower standardised RMSE) and higher precision (lower CV).*

Which parameters drive differences between group scans and focal follows?

We created a decision tree (Fig. 6), with simulations being classed as ‘same’ if the accuracy of the two approaches did not diverge by more than 1% (14% of cases); ‘focal better’ if focal follows outperformed the group scans (41% of cases); and ‘scan better’ when group scans outperformed the focal follows (45% of cases), representing situations where no prior information is available about the behaviour of interest or the study system. The most important factor determining which method performed better was behaviour duration: when behaviours were shorter than 23 seconds, focal follows clearly outperformed group scans, providing more accurate estimates in 78% of cases.

For behaviours longer than 23 seconds, the next most important parameter was behavioural visibility. Group scans performed better when behaviour visibility was high (at least 25% of behaviours that occur are observed), providing more accurate

estimates in 63% of those cases. When behavioural visibility was low (below 25% of actions observed), scans were generally better or the same as focal follows, depending on the action duration. Group scans performed better under high behaviour visibility (at least 25% of individuals are visible at any time), while focal follows performed better at very high behaviour visibility (over 75%) if actions were short (below 53 seconds).

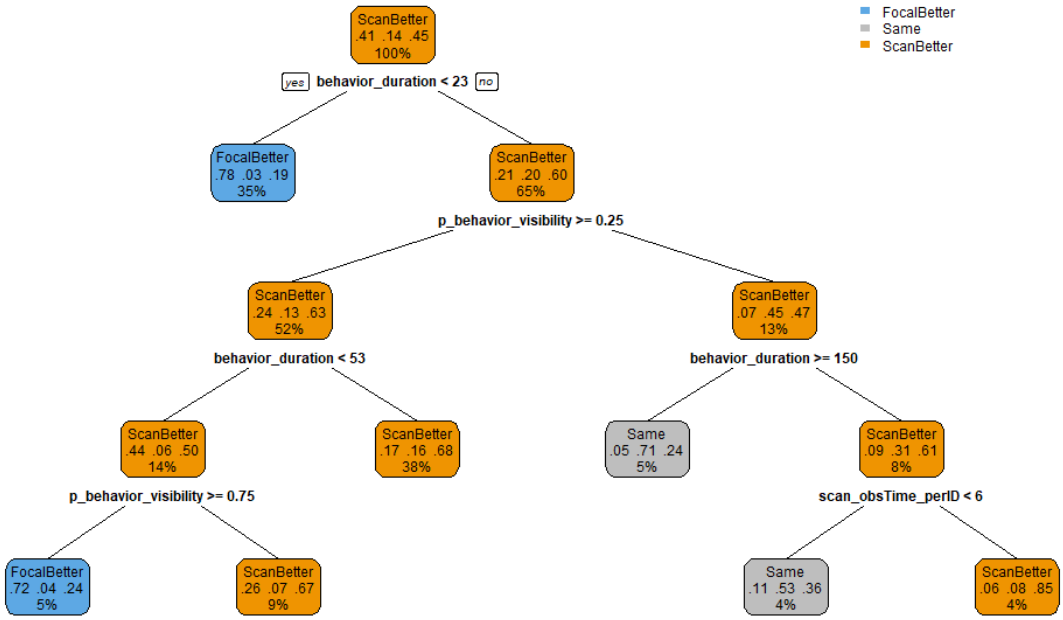

***Figure 6**. Decision tree showing whether focal follows (blue), group scans (yellow), or both methods (grey) provided the most accurate behavioural estimates. The tree should be read from top to bottom, with each split indicating a parameter and moving to the left meaning that the condition was fulfilled. Probabilities in the boxes represent the likelihood that focal follows, both methods, or group scans, respectively, performed best. Percentages indicate the proportion of all simulated cases falling into each category.*

Case Studies

1. ***What observation method should I use to record my behaviour of interest?***

For long, common, and highly visible behaviours like grooming or traveling, group scans (*scan interval time* = 5 minutes and s*can time per subject* = 3 seconds) generated accurate (Fig. 7A) and precise (Fig. 7B) behavioural estimates that were strongly correlated to the true values (Fig. 7C). Focal follows (*focal duration* = 15 minutes, *focal break time* = 1 minute) performed generally well, but generated estimates with lower accuracy, precision and correlation to the true values than group scans (Fig. 7).

For short, less common, and highly visible behaviours like aggression and drinking, both focal follows and group scans generated behavioural estimates of relatively low accuracy (Fig. 7A) and precision (Fig. 7B), but still a relatively high correlation to the true values (Fig. 7C). Estimates from focal follows were much more accurate than those from group scans, mainly because group scans tended to overestimate the true value (standardised RMSE > 100). The precision and correlation values were nearly indistinguishable between focal follows and group scans.

Finally, for short, less common, and less visible behaviours like threat postures and scratching, both focal follows and group scans generated behavioural estimates of relatively low accuracy (Fig. 7A), precision (Fig. 7B), and correlation to the true values (Fig. 7C). Estimates from group scans were more accurate than those from focal follows, because the low visibility of these behaviours prevented the overestimation that occurred in the more visible short behaviours. The precision and correlation values were nearly indistinguishable between focal follows and group scans.

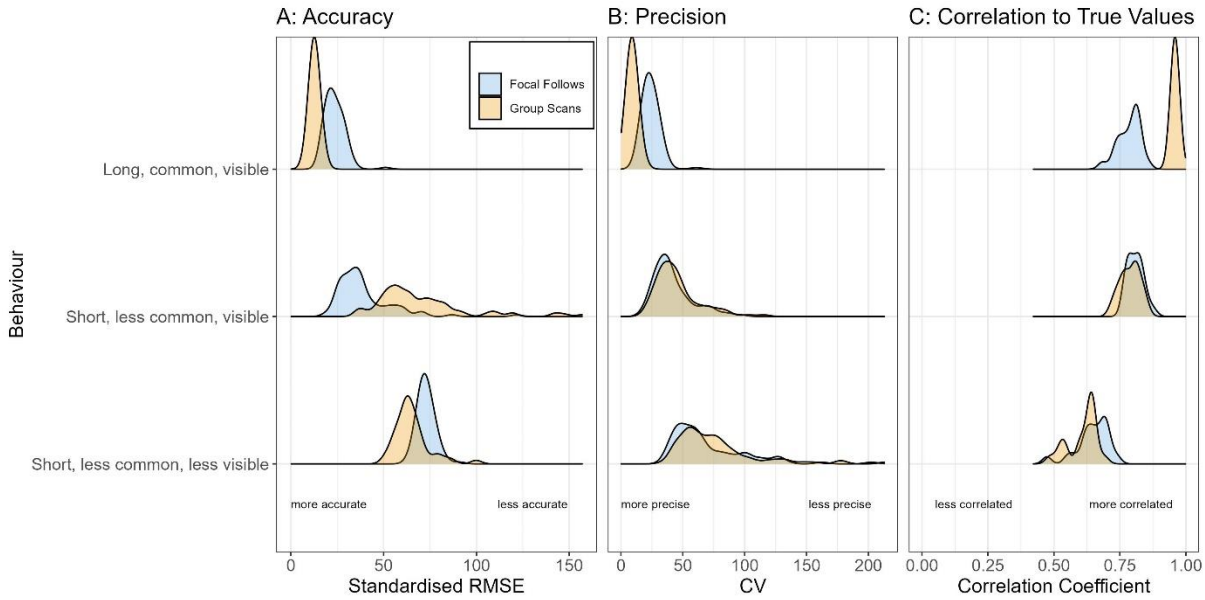

Figure 7: *Distributions of the A: accuracy, B: precision and C: correlation coefficient with the true value of behavioral estimates for a long, common and visible behaviour; short, rare, and visible behaviour; and short, rare, and less visible behaviour. Distributions in blue represent measures for behavioural estimates from focal follows, distributions in yellow represent measures for behavioural estimates from group scans. In plots A and B, measures closer to 0 indicate higher accuracy and precision, respectively, and therefore a better representation of the true underlying behavioural values. In plot C, measures closer to 1 indicate a higher correlation with the true value.*

2. *Is the relative performance of group scans versus focal follows influenced by group size?*

For long, common, and highly visible behaviours like grooming or traveling, group scans (*scan interval time* = 5 minutes and s*can time per subject* = 3 seconds) outperformed focal follows (*focal duration* = 15 minutes, *focal break time* = 1 minute) across all group sizes (Fig. 8A-C). Even for groups of 90 individuals, group scans provided estimates that were accurate, precise, and strongly correlated to the true value. Focal follows, in contrast, were strongly impacted by group size—for groups of 15 individuals, they performed nearly identical to group scans, but for 90 individuals, the group scans had a clear advantage.

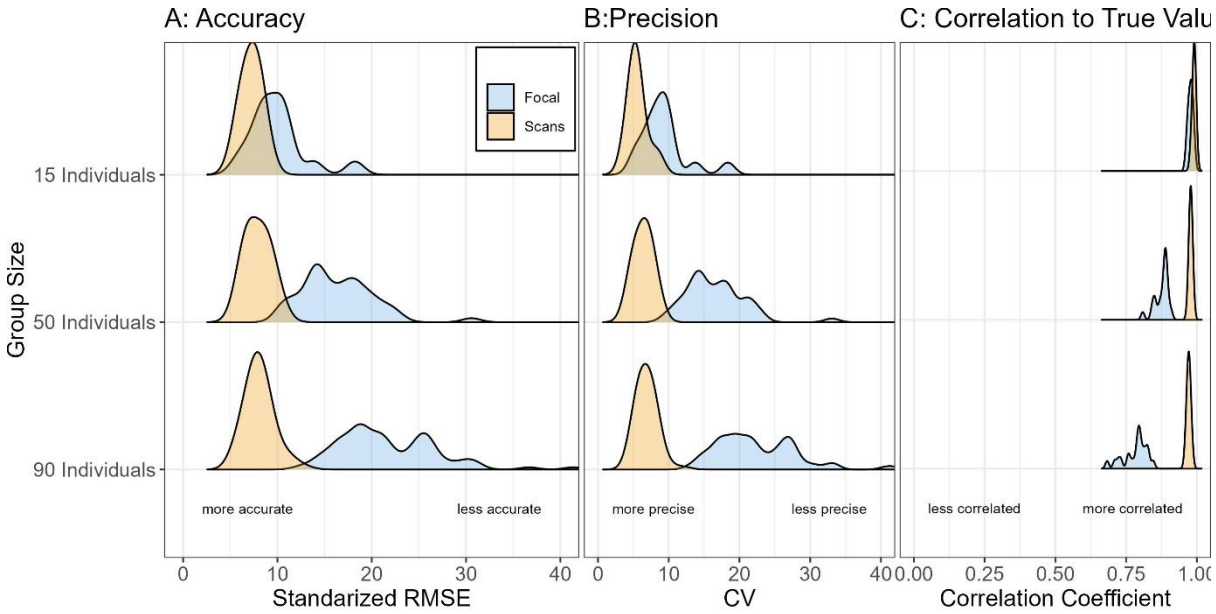

Figure 8*: Distributions of the A: accuracy, B: precision and C: correlation coefficient with the true value of behavioral estimates for groups of 15, 50, and 90 individuals. Distributions in blue represent measures for behavioural estimates from focal follows, distributions in yellow represent measures for behavioural estimates from group scans. In plots A and B, measures*

closer to 0 indicate higher accuracy and precision, respectively, and therefore a better representation of the true underlying behavioural values. In plot C, measures closer to 1 indicate a higher correlation with the true value.

3. How long does my study need to be?

For long, common, and highly visible behaviours like grooming or dustbathing, increasing the study duration improved the accuracy, precision and correlation to the true value of behavioural estimates from both focal follows and group scans (Fig. 9A-C). However, for group scans, the improvement was limited due to ceiling effects between 180 and 730 days. For focal follows, this ceiling was not yet reached by our maximum simulated study length of 730 days.

For short, less common, and highly visible behaviours like aggression, increasing the study duration also improved the accuracy, precision and correlation to the true value of behavioural estimates from both focal follows and group scans (Fig. 9D-E). This was especially true for group scans where the overestimation of values decreased with increasing study duration. Even at the longest study duration, the focal follows outperformed the group scans.

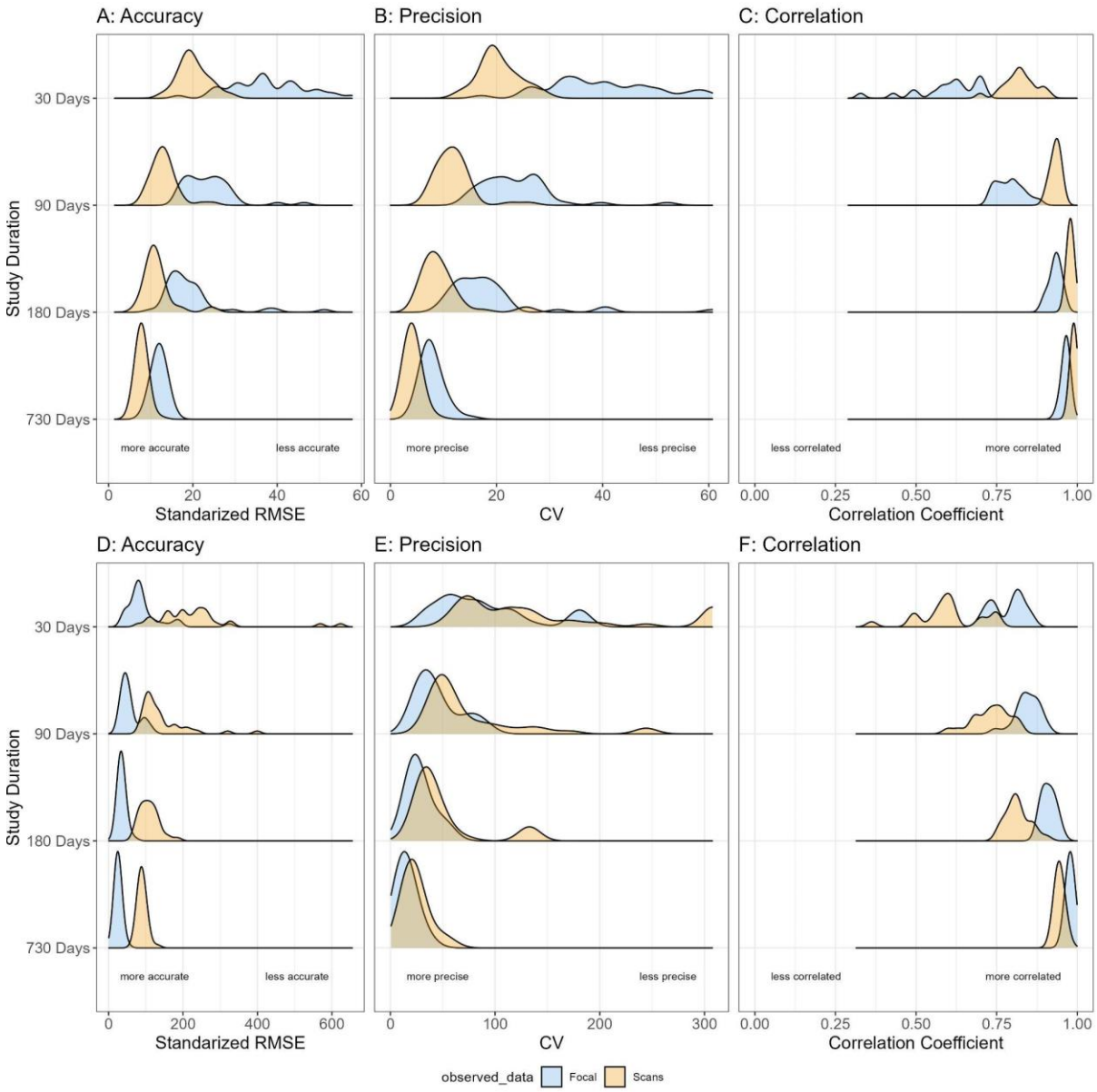

***Figure 9**: Distributions of the A/D: accuracy, B/E: precision and C/F: correlation coefficient with the true value of behavioral estimates for study durations 30, 90, 180, and 730 days for a long and common behaviour (A-C); and a short and rare behaviour (D-F). Distributions in blue represent measures for behavioural estimates from focal follows, distributions in yellow represent measures for behavioural estimates from group scans. In plots A and B, measures closer to 0 indicate higher accuracy and precision, respectively, and therefore a better representation of the true underlying behavioural values. In plot C, measures closer to 1 indicate a higher correlation with the true value.*

Discussion

In this simulation study, we evaluated the performance of two commonly used observation methods in animal behavioural research—focal follows and group

scans—in estimating the true occurrence of behaviours. By comparing the behavioural estimates generated by each method to the known ground truth in our simulations, we were able to assess their accuracy (how close estimates are to the true value), precision (how consistent estimates are across repeated samples) and correlation to the true value. This approach provides a quantitative foundation for guiding researchers in selecting the most appropriate observational strategy depending on the study system, the behaviour of interest, and the observational setup. It also offers insight into how comparable behavioural data are when they were collected using different methods.

Our first step was to evaluate how focal follows and group scans perform across a broad range of conditions. As expected, both methods produced behavioural estimates that, on average, showed relatively low accuracy, precision, and correlation with the true behavioural occurrences. This reflects a well-recognized reality in behavioural research: observational data almost always represent only a sample of the full set of behavioural occurrences, and some level of discrepancy is inevitable (Altmann, 1974; De Moor et al., 2024). The central insight of this approach was that these discrepancies differ systematically between the two observation methods. Focal follows tended to underestimate the true proportion of time each individual spent engaged in a behaviour, because they only recorded behaviours that took place while an individual was the subject of a focal follow, missing many events that occurred outside of that window. Group scans, on the other hand, tended to overestimate the proportion of time spent engaged in a behaviour, because any behaviour observed during a scan was considered to have occurred throughout the entire interval—even if it only happened momentarily. As a result, behaviours, especially rare or brief ones, could appear to occupy a larger proportion of time than they actually did. This systematic difference between both observation methods was also reflected in the generally low correlation between behavioural estimates from focal follows and group scans of the same true behavioural occurrences. These findings highlight the need for caution when interpreting and comparing behavioural estimates derived from different observation methods.

When evaluating system, behaviour, and observation parameter effects on method performance, we found that some parameters affected both focal follows and group

scans in similar ways: longer *study duration,* higher *behaviour frequency* and better *behaviour visibility* improved the accuracy, precision, and correlation of behavioural estimates with true behavioural occurrences. Other parameters had contrasting effects on the two methods. *Group size* impacted focal follow performance more strongly, whereas *terrain visibility* (i.e. the proportion of the group that was visible) mattered more for group scans. This is because, for focal follows, the observation time per individual decreases as group size increases, reducing the density of data collected for each subject. For group scans, in contrast, performance remained high even in large groups, as many individuals could be sampled reliably at each time point. However, this advantage quickly disappeared when the terrain visibility was low, especially when scans were infrequent. An empirical study of arboreal wedge-capped capuchins (*Cebus oliuaceus*) illustrates this effect: behaviours that occurred in the middle of the canopy, where visibility is lowest, were less likely to be recorded using group scans than focal follows (Fragaszy et al., 1992). The balance between group size and terrain visibility is therefore a key factor in determining which method performs better under different conditions.

Behaviour duration had a much stronger impact on group scans' performance than on focal follows. While focal follows were equally likely to detect short and long behaviours, short behaviours were more easily missed when using group scans, as they could occur between sampling points. In addition, the time spent engaged in short behaviours was more likely to be overestimated when using group scans, since a brief behaviour recorded during a scan was considered to have occurred throughout the entire interval (Altmann, 1974; Brereton et al., 2022; Griffin & Adams, 1983). These patterns were also reflected in the positive effect of *scan interval time* (and, to a lesser extent, *scan time per subject*): more frequent scans improved the accuracy and precision of group scan estimates by reducing both the likelihood of missing behaviours and the extent of overestimation. This aligns with findings from a study on howler monkeys (*Alouatta pigra*), which showed that longer behaviours were accurately estimated when sampled using group scans with short intervals. Shorter behaviours were more likely to be missed, but this was partly balanced by an overestimation of their duration, leading to only moderate overall error (Amato et al., 2013). A study on lambs (*Ovis aries*) and another on cows (*Bos taurus*) also found

that increasing the frequency of scans improved the accuracy of detecting shorter behaviours (Hämäläinen et al., 2016; Pullin et al., 2017).

Finally, *focal duration* and *focal break time* had no measurable impact on performance in our simulations. However, this may be an artefact of our modelling assumptions, which treated behavioural occurrences as independent of each other. In reality, behaviours may cluster in time (e.g., because individuals might perform the same action repeatedly once they start, or because certain behaviours tend to happen most at certain times of the day). In such cases, longer focal durations could lead to some individuals being observed during active periods and others during inactive ones. This can increase temporal sampling bias and reduce the representativeness of behavioural estimates compared to group scans, which distribute sampling more evenly across individuals and time. However, the impact of such temporal bias was found to be minimal in an empirical study of wedge-capped capuchins, where variation across sampling days had no significant effect on estimates of activity budgets, suggesting that temporal clustering may not always introduce substantial bias in behavioural estimates from focal follows (Fragaszy et al., 1992).

Overall, focal follows tended to outperform group scans for short, less common behaviours. This is in line with foundational guidance on observation methods (Altmann, 1974; Bateson & Martin, 2021), which emphasized two key advantages of focal follows: their ability to capture short or rare behaviours, and their ability to record the temporal sequence of events. In addition, focal follows allow to estimate both the rate at which behaviours occur, as well as their average duration, whereas group scans only allow to estimate the probability of an individual engaging in a behaviour (although approaches to estimate frequencies and duration from scan data exist, e.g. Griffin & Adams, 1983; Suen & Ary, 1984). These features established focal follows as the method of choice in many behavioural studies (Brereton et al., 2022; Webber & Vander Wal, 2019). However, our simulations show that for longer and more common behaviours, group scans consistently outperformed focal follows—particularly in larger groups with relatively good terrain visibility and in shorter studies. This is because group scans allow for the near-simultaneous sampling of many individuals, increasing the density of behavioural data that is recorded. When behaviours are long and common, this increased sampling density outweighs the

limitation that some occurrences may be missed between scan intervals. This pattern is consistent with an empirical study on olive baboons (*Papio anubis*), in which behavioural records derived from GPS collars were subsetted to mimic sampling via focal follows and group scans. The study found that group scans produced more accurate estimates of spatial association—a long and relatively common behaviour (Davis et al., 2018).

Many behaviours of central interest to researchers—including feeding, travelling, resting, grooming, and spatial association—tend to be relatively long and common. For these behaviours, group scans conducted at short intervals can be an efficient method for data collection. At the same time, other behaviours of interest—such as vigilance, agonistic interactions and sexual behaviours—are typically shorter and may be better captured through focal follows. For instance, three studies on howler monkeys, white-faced capuchins (*Cebus capucinus*), and lambs respectively found that group scans generated accurate estimates of activity budgets while needing lower observation effort than focal follows. However, for less common and shorter behaviours, only focal follows generated accurate estimates (Amato et al., 2013; Pullin et al., 2017; Rose, 2000). We therefore suggest, as others have (Canteloup et al., 2020; Fragaszy et al., 1992; Rose, 2000), that a mixed approach may often be most effective. For example, researchers could combine focal follows with periodic scans of all visible individuals (e.g., every five minutes) or conduct small-group focal follows to increase sampling coverage while retaining sequential detail (e.g. Dragić et al., 2022). In some cases, combining systematic sampling methods with *ad libitum* observations may also be appropriate—particularly when *ad libitum* data correlate well with systematically sampled data—in which case the benefits of increased data density might outweigh the potential costs of sampling bias (Archie et al., 2014; Canteloup et al., 2020). However, it is important to note that incorporating *ad libitum* data complicates the estimation of observation effort, a factor that should be carefully considered when choosing an observation approach (Milinsky, 1997). Ultimately, hybrid strategies allow researchers to capitalize on the complementary strengths of each method and therefore improve the overall quality of their behavioural data samples.

We implemented a simulation-based approach in this study because it allowed us to set a ground truth against which to evaluate the performance of different observation methods. However, simulations inevitably simplify complex systems and must be interpreted with care (Fogarty et al., 2022; Williams et al., 2024). While our simulated groups were parameterised to resemble real-world conditions, they did not account for factors such as behavioural autocorrelation, changes in visibility, or dynamic group composition. As such, our findings can only ever be a general guide and need to be complemented by empirical validation and system-specific knowledge. Our simulations are also not meant to find the 'best' observation method, but rather to provide guidance on what variables to consider when choosing an observational approach. We encourage readers to use the SIMBO app to explore how different methods perform under conditions relevant to their specific system and research question.

Recognizing the strengths and limitations of focal follows and group scans is essential—not only for guiding the choice of observation method for a given study, but also for interpreting behavioural estimates more broadly. Behavioural estimates always represent only a subset of the true behavioural occurrences, and different methods can introduce different sampling biases, such that estimates of the same behaviour can differ. Careful consideration of how well behavioural observations reflect the true behavioural occurrences of interest can guide sound methodological choices, inform modelling approaches to account for sampling biases, and enable more accurate comparisons across methods and studies—ultimately improving our understanding of animal behaviour.

Data Availability: Scripts and results can be found here: https://github.com/AlexMielke1988/Observation_Methods_Comparison

Acknowledgements

D.D.M. and L.J.N.B. acknowledge funding from a European Research Council Consolidator Grant (FriendOrigins - 864461). DDM acknowledges funding from the Max Planck Society and a British Academy/Leverhulme Small Research Grant

(SRG23\231253). L.J.N.B acknowledges funding from the National Institutes of Health (R01AG087902, R01AG084706, R01AG060931, R01MH118203).

References

Aldrich-Blake, F. P. G. (1970). Problems of social structure in forest monkeys. In J. H. Crook (Ed.), Social behaviour of birds and mammals (pp. 79-101). Academic Press.

Altmann, J. (1974). Observational study of behavior: Sampling methods. Behaviour, 49(3-4), 227-266. https://doi.org/10.1163/156853974X00534

Amato, K. R., Van Belle, S., & Wilkinson, B. (2013). A comparison of scan and focal sampling for the description of wild primate activity, diet and intragroup spatial relationships. Folia Primatologica, 84(2), 87-101. https://doi.org/10.1159/000348305

Archie, E. A., Tung, J., Clark, M., Altmann, J., & Alberts, S. C. (2014). Social affiliation matters: Both same-sex and opposite-sex relationships predict survival in wild female baboons. Proceedings of the Royal Society B, 281(1793), 20141261. https://doi.org/10.1098/rspb.2014.1261

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67, 1-48.

Bateson, M., & Martin, P. (2021). Measuring behaviour: An introductory guide. Cambridge University Press.

Brereton, J. E., Tuke, J., & Fernandez, E. J. (2022). A simulated comparison of behavioural observation sampling methods. Scientific Reports, 12(1), 3096. https://doi.org/10.1038/s41598-022-07169-5

Canteloup, C., Puga-Gonzalez, I., Sueur, C., & van de Waal, E. (2020). The effects of data collection and observation methods on uncertainty of social networks in wild primates. American Journal of Primatology, 82(7), e23137. https://doi.org/10.1002/ajp.23137

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. https://doi.org/10.1016/j.anbehav.2014.07.023

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. https://doi.org/10.1016/j.anbehav.2018.04.012

De Moor, D., Brent, L. J. N., Silk, M. J., & Brask, J. B. (2024). Layers of latency in social networks and their implications for comparative analyses. EcoEvoRxiv. https://doi.org/10.32942/X2G894

Dragić, N., Keynan, O., & Ilany, A. (2022). Protocol to record multiple interaction types in small social groups of birds. STAR Protocols, 3(4), 101814. https://doi.org/10.1016/j.xpro.2022.101814

Fogarty, L., Ammar, M., Holding, T., Powell, A., & Kandler, A. (2022). Ten simple rules for principled simulation modelling. PLoS Computational Biology, 18(3), e1009917. https://doi.org/10.1371/journal.pcbi.1009917

Fragaszy, D. M., Boinski, S., & Whipple, J. (1992). Behavioral sampling in the field: Comparison of individual and group sampling methods. American Journal of Primatology, 26(4), 259-275. https://doi.org/10.1002/ajp.1350260404

Gilby, I. C., Pokempner, A. A., & Wrangham, R. W. (2010). A direct comparison of scan and focal sampling methods for measuring wild chimpanzee feeding behaviour. Folia Primatologica, 81(5), 254-264. https://doi.org/10.1159/000322354

Griffin, B., & Adams, R. (1983). A parametric model for estimating prevalence, incidence, and mean bout duration from point sampling. American Journal of Primatology, 4(3), 261-271. https://doi.org/10.1002/ajp.1350040305

Hämäläinen, W., Ruuska, S., Kokkonen, T., Orkola, S., & Mononen, J. (2016). Measuring behaviour accurately with instantaneous sampling: A new tool for selecting appropriate sampling intervals. Applied Animal Behaviour Science, 180, 166-173. https://doi.org/10.1016/j.applanim.2016.04.006

Hepworth, G., & Hamilton, A. (2001). Scan sampling and waterfowl activity budget studies: Design and analysis considerations. Behaviour, 138(11-12), 1391-1405. https://doi.org/10.1163/156853901317367654

Ihle, M., Winney, I. S., Krystalli, A., & Croucher, M. (2017). Striving for transparent and credible research: Practical guidelines for behavioral ecologists. Behavioral Ecology, 28(2), 348-354. https://doi.org/10.1093/beheco/arx003

Kappeler, P. M., Barrett, L., Blumstein, D. T., & Clutton-Brock, T. (2013). Constraints and flexibility in mammalian social behaviour: Introduction and synthesis. Philosophical Transactions of the Royal Society B, 368, 20120337. https://doi.org/10.1098/rstb.2012.0337

Karniski, C., Patterson, E. M., Krzyszczyk, E., Foroughirad, V., Stanton, M. A., & Mann, J. (2014). A comparison of survey and focal follow methods for estimating individual activity budgets of cetaceans. Marine Mammal Science, 31(3), 839-852. https://doi.org/10.1111/mms.12198

Lehner, P. N. (1998). Handbook of ethological methods. Cambridge University Press.

Leimar, O., Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2022). Behavioural specialization and learning in social networks. Proceedings of the Royal Society B, 289(1980), 20220954. https://doi.org/10.1098/rspb.2022.0954

Lukas, D., & Clutton-Brock, T. (2017). Comparative studies need to rely both on sound natural history data and on excellent statistical analysis. Royal Society Open Science, 4(11), 171211. https://doi.org/10.1098/rsos.171211

Mielke, A., Preis, A., Samuni, L., Gogarten, J. F., Lester, J. D., Crockford, C., & Wittig, R. M. (2021). Consistency of social interactions in sooty mangabeys and chimpanzees. Frontiers in Ecology and Evolution, 8, 603677. https://doi.org/10.3389/fevo.2020.603677

Milinsky, M. (1997). How to avoid seven deadly sins in the study of behavior. Advances in the Study of Behaviour, 26, 159-180. https://doi.org/10.1016/s0065-3454(08)60379-4

Nunn, C. L. (2011). 11. Behavior, Ecology, and Conservation of Biological and Cultural Diversity. In The Comparative Approach in Evolutionary Anthropology and Biology (pp. 255-279). University of Chicago Press. https://doi.org/10.7208/9780226090009-012

O'Dea, R. E., Parker, T. H., Chee, Y. E., Culina, A., Drobniak, S. M., Duncan, D. H., Fidler, F., Gould, E., Ihle, M., Kelly, C. D., Lagisz, M., Roche, D. G., Sanchez-Tojar, A., Wilkinson, D. P., Wintle, B. C., & Nakagawa, S. (2021). Towards open, reliable, and transparent ecology and evolutionary biology. BMC Biology, 19(1), 68. https://doi.org/10.1186/s12915-021-01006-3

Pullin, A. N., Pairis-Garcia, M. D., Campbell, B. J., Campler, M. R., & Proudfoot, K. L. (2017). Technical note: Instantaneous sampling intervals validated from continuous video observation for behavioral recording of feedlot lambs. Journal of Animal Science, 95(11), 4703-4707. https://doi.org/10.2527/jas2017.1835

Rose, L. M. (2000). Behavioral sampling in the field: Continuous focal versus focal interval sampling. Behaviour, 137(2), 153-180. https://doi.org/10.1163/156853900502006

Rubenstein, D. R., & Abbot, P. (2017). Comparative Social Evolution. Cambridge University Press. https://doi.org/10.1017/9781107338319

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution, 1(2), 103-113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Suen, H. K., & Ary, D. (1984). Variables influencing one-zero and instantaneous time sampling outcomes. Primates, 25(1), 89-94. https://doi.org/10.1007/BF02382298

Therneau, T., Atkinson, B., & Ripley, B. (2025). Rpart (Version 4.1.24). In

Webber, Q. M. R., & Vander Wal, E. (2019). Trends and perspectives on the use of animal social network analysis in behavioural ecology: A bibliometric approach. Animal Behaviour, 149, 77-87. https://doi.org/10.1016/j.anbehav.2019.01.010

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? Nature, 582, 337–340. https://doi.org/10.1038/d41586-020-01751-5

Whitehead, H. (2008). Analyzing Animal Societies - Quantitative Methods for Vertebrate Social Analysis. University of Chicago Press.

Williams, C., Yang, Y., Lagisz, M., Morrison, K., Ricolfi, L., Warton, D. I., & Nakagawa, S. (2024). Transparent reporting items for simulation studies evaluating statistical methods: Foundations for reproducibility and reliability. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210x.14415