International Zoo News Vol. 50, No. 5 (2003), pp. 288-297

VARIATION IN RELIABILITY OF MEASURING BEHAVIOURS OF REINTRODUCED ORPHAN GORILLAS

BY TONY KING, ELKE BOYEN AND SANDER MUILERMAN

Introduction

In any study of animal behaviour, it is important (although often overlooked) to test the reliability of the data collected (Martin and Bateson, 1993). Such reliability-testing allows quantification of the level of confidence which can be given to the interpretation of the results. Without this testing, it cannot be determined whether the results provide a true reflection of the behaviours performed by the study animals, or are simply a function of the methodology employed, or of the observers themselves. Therefore, this study aimed to investigate the reliability of data collected during a quantitative survey of the behaviour of a reintroduced group of orphan gorillas in the Republic of Congo.

Project background and study group

Despite the long history of behavioural and ecological research on the mountain gorilla (*Gorilla beringei beringei*) (e.g. Schaller, 1963), the much more widespread western lowland gorilla (*Gorilla gorilla gorilla*) is still relatively unknown in the wild (Dixson, 1981). This imbalance may be attributed to the successful habituation of several groups of mountain gorillas, something that is proving difficult to achieve in the wild with the seemingly more elusive western lowland gorilla. However, an ambitious programme run by the John Aspinall Foundation, a UK-based charitable organisation, in partnership with the governments of the Republic of Congo and of Gabon, provides a unique opportunity to study western lowland gorillas at close-range in their natural habitat. The Projet Protection des Gorilles has been working in Congo since 1986, and in Gabon since 1998, facilitating the confiscation by the respective governments of young western lowland gorillas orphaned by the bush-meat trade, rehabilitating them, and eventually reintroducing them in small groups to protected areas from which the species has been extirpated in recent history (Courage *et al.*, 2001; Cousins, 2002).

These reintroduced groups are, by necessity, of artificial composition. They are relatively even-aged, unrelated, of random sex-ratios, and, at least upon release, with all members still considered juvenile or sub-adult. In contrast, naturally-occurring gorilla groups in West Africa are generally composed of a single adult ‘silver-back’ dominant male, plus several adult females, sub-adult ‘black-back’ males, juveniles and infants (Parnell, 2002). However, they have been raised in close-contact with humans, and so can now be relatively easily observed as they grow and mature.

The group of released gorillas selected for the initial behavioural survey, and therefore for this reliability study, consisted of ten (4.6) individuals, aged between 2.5 and 8 years (Table 1). Nine of these had been fully released in June 2001 into the Reserve Naturelle des Gorilles de Lesio-Louna, Republic of Congo, an area of extensive *Loudetia* and *Hyparrhenia* grasslands with swampy gallery forests along watercourses (see Courage *et al.*, 2001; Cousins, 2002). The tenth and youngest (Hélène) was introduced to the group in July 2001 before finally being released with them in November 2001. The group had lived in complete freedom since the full release, although with almost daily monitoring by local staff, and were nutritionally independent except for the giving of milk after the introduction of Hélène.

Table 1. Composition of gorillas within the reintroduced group used for this study.

Code	Name	Sex	Estimated age (years)	Arrival date
F1	Koto	female	8	1996
M1	Djeke	male	7.5	1996
M2	Kelle	male	6.5	1996
M3	Pikounda	male	5.5	1999
F2	Mpoumbou	female	5.5	1997
F3	Massabi	female	5	1998
F4	Tchivoulou	female	5	1999
M4	Kama	male	4	1998
F5	Louboko	female	3.5	1999
F6	Hélène	female	2.5	2001

Methods

The methodology for measuring behaviour of the group was developed and tested during July 2002. Focal animal sampling, i.e. recording information for just a single individual during a sample period, was found to be the most practical sampling rule, due to the combination of the tendency of several members of the group to disperse while foraging and the low-range of visibility within the forest. It was necessary to utilise two separate recording methods simultaneously to adequately sample the wide-range of behaviours performed by the gorillas. Common or lengthy behaviours were most easily sampled by instantaneous recording (noting the behaviour being performed at regular instants in time). However, rare or short-duration behaviours could only be adequately sampled by continuous recording (noting the occurrence of a behaviour every time it is performed during a period of time).

Therefore, a list of 65 behaviours was compiled following the preliminary observations, each defined carefully to ensure a high level of mutual exclusivity, and to avoid differences in interpretation between observers or over time. Each behaviour was then assigned to the appropriate recording methodology, with those assigned to continuous recording being referred to as ‘target’ behaviours. Some rare but lengthy behaviours were included in both recording methods. It was envisaged that the two recording methods would be analysed separately, with the instantaneous recording giving proportions of time spent in common or lengthy behaviours, and the continuous recording giving frequencies (but not durations) of rare or short-duration (target) behaviours.

Each sample period was set at 50 minutes. Instantaneous data was taken at 1-minute intervals, for 50 sample points. On each sample point, the height from the ground of the focal gorilla was recorded (in metres), the behaviour of the focal gorilla, the direction of the behaviour (if relevant), the name of the gorilla closest to the focal gorilla (if the behaviour of the focal gorilla was directional, the name of the gorilla involved in the behaviour of the focal gorilla was recorded even if it was not in fact the closest), the distance (in metres) to that gorilla (0 was recorded if the gorillas were in physical contact), and if feeding, the food-type (eg leaf, fruit, shoot, faeces), and the name of the food (if known).

Continuous data was taken simultaneously throughout the sample period, between each sample point and to the end of the fiftieth minute. All occurrences of target behaviours performed by the focal gorilla were recorded. Target behaviours performed by non-focal gorillas were also recorded if they were considered to involve or impact the focal gorilla, with the name of the gorilla exhibiting the behaviour, and, if relevant, the direction and recipient of

the behaviour. Additional non-target behaviours were also recorded if considered to contribute to a sequence of target behaviours.

Time of day was split into three periods (late morning: 08:30-10:59; midday: 11:00-12:59; early afternoon:13:00-15:29). Each sample period was fitted entirely into just one of these time-periods. For this reliability study, one sample was taken in each of the three time periods, each day for three consecutive days (28-30/08/2002), giving a total of nine sample periods. The youngest gorilla was excluded from the sampling, while the remaining nine gorillas were sampled once each, with the sampling order selected at random. For the purposes of the reliability-testing, each sample was recorded simultaneously by three observers, except one sample which was missed by one of the observers.

Instantaneous data was summarised for each sample period for each observer as proportions of sample points spent by the focal gorilla in performing each particular behaviour, and in proximity to each other individual gorilla. In addition to the 1-minute interval analysis, the instantaneous data was also analysed as if collected at 5- or 10-minute intervals, to investigate the effect of sample interval on the results and their reliability.

The eldest (Koto) and youngest (Hélène) members of the study group at rest, a common behaviour of lengthy duration best measured by instantaneous sampling. (Photo: Tony King)

Continuous data was summarised for each sample period for each observer as frequency of occurrence of each target behaviour within the sample period, both by the focal gorilla and by non-focal gorillas.

The reliability of the results obtained for each behaviour was quantified by comparison of these proportions and frequencies between each of the three pairs of observers. Pearson correlation coefficients (r) were calculated between each observer-pair, for each specific behaviour. Coefficient values approaching 1 indicate high inter-observer reliability, and therefore provide a high level of confidence in the validity of the results and therefore the recording methods. Values approaching 0, and negative correlations, indicate low inter-observer reliability, and therefore highlight unreliable results and inadequacies in the recording methods for that specific behaviour. To summarise several values, average Pearson correlation coefficients were calculated by converting each coefficient to their Fisher *z* transforms ($z = 0.5\ln[(1+r)/(1-r)] = \tanh^{-1}r$), calculating the arithmetic mean of the *z* transforms, and converting the mean *z* transform back into a correlation (Martin and Bateson 1993). However, the *z* transform was invalid for correlations of 1, so median values were also used to summarise several correlations.

Results

Fifty-three behaviours were recorded during this trial study, 49 from the original list of 65 defined following the preliminary observations, plus an extra four. Two of the extra behaviours were added to provide information when the focal gorilla was out-of-sight, but when its location and general activity was known (out-of-sight: rest/forage and out-of-sight:

aggressive interaction). However, many of the behaviours were recorded only very rarely, and so for ease of presentation and comprehension, the following analysis has concentrated only on those behaviours which made up 90% of the data collected. In all observer-comparison tables, sample sizes (n) are nine for observer pair 1 and eight for observer pairs 2 and 3.

During the course of the nine sample-periods of instantaneous recording, the three observers recorded a total of 27 behaviours, although 90% of the data recorded was accounted for by just eight behaviours, and 26% by a single behaviour (Table 2). For six of these eight behaviours, inter-observer reliability was very high, with Pearson Correlation Coefficients of over 0.90 for all observer-comparisons. For the other two behaviours, inter-observer reliability was fairly high for one (Observe, average r = 0.91), but low for the other (Move, average r = 0.37).

The closest-neighbour data exhibited a similarly high level of inter-observer reliability, with only the least-frequently recorded closest-neighbour showing an average correlation below 0.90 (Table 3). There were also very high-levels of inter-observer reliability for the recording of the major food-types eaten, although the occurrence of ant-eating was over-looked by two of the three observers (Table 4).

Increasing the sample-interval decreased the total number of behaviours recorded, but had little impact on the summary proportions of the commonly-observed behaviours, with the exception of 'Eat' which decreased with increasing the sample-interval. However, it did affect inter-observer reliability, which decreased with an increase in the sample-interval (Table 5).

During the continuous recording, the three observers recorded a total of 18 target behaviours for the focal gorilla, with a further 19 non-target behaviours recorded to supplement the understanding of the target behaviour sequences. Twenty target behaviours were also recorded for non-focal gorillas, as they were believed by the observers to impact the focal gorilla, with an additional 15 behaviours recorded as supplementary information. The inter-observer reliability for the continuous recording was generally lower than for the instantaneous recording, although high correlations were achieved for some of the more frequent and unambiguous target behaviours (eg 'Chest-beat' and 'Charge') (Tables 6 and 7). Blank values indicate behaviours that were overlooked by one or both observers (eg 'Hit' was recorded by two of the three observers, and 'Posture' by just one observer).

Discussion

This study has highlighted variations in the reliability of measuring different behaviours of a group of orphan gorillas reintroduced to their natural habitat, variations that need to be considered when interpreting results from future behavioural studies using these or similar methodologies. Most of the behaviours measured by instantaneous sampling, that is behaviours that are commonly performed or of lengthy duration, and other variables such as nearest-neighbour and food-parts consumed, were shown to be very reliably measured. One notable exception was the unreliable measurement of a commonly performed behaviour, 'Move'. Variation in reliability was higher for behaviours recorded by continuous sampling, that is behaviours rarely performed or of short-duration. Only one of these behaviours was measured as reliably as those recorded by instantaneous sampling. Several others were measured less reliably, but to a level that should be considered acceptable to give a reasonable impression of the behaviours performed in reality by the focal gorilla. The reliability of measuring behaviours of non-focal gorillas was lower again, although despite the increased influence of observer-interpretation in deciding what to record, a selection of behaviours were

measured to what should be considered an acceptable level of reliability, particularly those involving direct aggressive interaction with the focal gorilla.

The dominant male (Djeke) mounting the dominant female (Koto), a rare behaviour best measured by continuous sampling. (Photo: Elke Boyen and Sander Muilerman)

The second-ranked male (Kelle) exhibiting at least three separate behaviours (vocalisation, chest-beat and charge), a frequent combination perhaps best analysed as a single behavioural sequence (e.g. 'chest-beat display'). (Photo: Elke Boyen and Sander Muilerman)

The variable reliability scores presented here illustrate how care must be taken when interpreting results from behavioural studies in general, and the proposed studies of the behaviour of the released groups of orphan gorillas within the Projet Protection des Gorilles. However, they can also be used to identify modifications to data-collection and analysis that would increase the reliability of future results. While most common behaviours can be measured by the methods described here, it may be necessary to merge several similar behaviours into more general categories for a truer representation of general activity patterns. For example, behaviours such as 'Climb' and 'Descend' could be incorporated into the more general 'Move', at least for analysis purposes. The advantage of retaining these more precise descriptions during the data-collection phase is that it allows reclassification at a later date should the need arise. This would be particularly important should the results eventually be compared with other studies that use similar but not identical methods. Similarly, some of the less-reliable target behaviours may need to be grouped into behavioural sequences, rather than analysed separately, as they often occur dependently of each other (eg a 'Chest-beat' may often be followed by a 'Charge', a 'Smash vegetation' and a 'Posture', a display sequence that would generally be noted by any observer, even if different observers may miss or interpret differently one or more of the components). Such grouping of behaviours into sequences should increase the reliability of the data, and should allow more meaningful comparison between observers and over time. Again, though, it is worth retaining the increased level of detail at the data-collection phase.

The sample interval analysis illustrated how reliability declined with increasing sample interval. Therefore the 1-minute sample interval will be retained for the proposed behavioural study of this group. However, the results from the 5-minute sample interval were also mostly highly reliable, and the use of this sample interval may be beneficiary in allowing extra time in the field for making observations of rare or discrete behaviours, and in reducing data-input requirements.

In conclusion, this study has highlighted the variability in the reliability of measuring different behaviours, a factor that is often overlooked in behavioural studies despite it's

implications for the interpretation of the results. Never-the-less, the reliability of the data collected in this study was high for the majority of behaviours and related variables. Therefore the methods described here, with some minor modifications, should provide reliable quantitative data on the behaviour of this reintroduced group of orphan gorillas, and other similar groups within the Projet Protection des Gorilles, and should be repeatable by different observers for comparisons over time and between groups. We hope that long-term measuring of behaviour will provide the opportunity to assess the development and adaptation of the reintroduced orphan gorillas, and to develop and test hypotheses on the behaviour and ecology of the western lowland gorilla in its natural habitat.

Acknowledgements
We thank the Ministère de l'Economie Forestière et de l'Environnement of the Republic of Congo, and the John Aspinall Foundation of UK, for permission to undertake this research in the Lesio-Louna Reserve. We also thank the staff of the Projet Protection des Gorilles, Republic of Congo, in particular Ian Henderson, Christelle Chamberlan, Nicaise Ngoulou, Ossanda, Ghislain Mvila and Albertine Ndokila. Waterproof notebooks were generously donated by BCB International Ltd.

References
Courage, A., Henderson, I. and Watkin, J. (2001): Orphan gorilla reintroduction: Lesio-Louna and Mpassa. *Gorilla Journal* 22: 33-35.
Cousins, D. (2002): Natural plant foods utilized by gorillas in the former Brazzaville Orphanage and the Lesio-Louna Reserve. *International Zoo News* 49(4): 210-218.
Dixson, A. F. (1981): *The Natural History of the Gorilla*. Weidenfeld and Nicolson, London.
Martin, P. and Bateson, P. (1993): *Measuring Behaviour: An Introductory Guide*. Second Edition. Cambridge University Press.
Parnell, R. J. (2002): Group size and structure in Western Lowland Gorillas (Gorilla gorilla gorilla) at Mbeli Bai, Republic of Congo. *American Journal of Primatology* 56:193-206.
Schaller, G. B. (1963): *The Mountain Gorilla: Ecology and Behaviour*. University of Chicago Press, Chicago.

Tony King, Projet Protection des Gorilles, BP 13977, Brazzaville, Republic of Congo, Elke Boyen and Sander Muilerman, Hooigracht 69, 2312 KP Leiden, The Netherlands.

Table 2. Inter-observer reliability, given as Pearson Correlation Co-efficients, for behaviours* conducted by the focal gorilla during instantaneous recording at 1-minute intervals. (*8 most frequently recorded behaviours (of 27 recorded), accounting for over 90% (91.3%) of data recorded).

Behaviour	**Proportion**	**Observer pair 1**	**Observer pair 2**	**Observer pair 3**	**Average**
Rest	0.26	0.91	0.94	0.98	0,95
Eat	0.18	0.95	0.98	0.98	0.97
Out-of-sight - rest/forage	0.14	0.98	0.99	1.00	0.99
Play	0.12	0.91	0.94	0.99	0.96
Sleep	0.10	0.99	0.99	1.00	1.00
Observe	0.06	0.98	0.87	0.74	0.91
Move	0.04	-0.16	0.80	0.22	0.37
Self-play	0.02	1.00	0.99	0.99	0.99
Average		0.94	0.97	0.98	0.96

Table 3. Inter-observer reliability, given as Pearson Correlation Co-efficients, for closest-neighbour data collected during 1-minute instantaneous recording.

Neighbour code	Proportion	Observer pair 1	Observer pair 2	Observer pair 3	Average
F1	0.21	0.99	1.00	1.00	1.00
M2	0.14	0.98	0.89	0.93	0.94
F4	0.13	0.99	0.98	0.99	0.99
M1	0.10	0.89	0.92	0.90	0.90
F2	0.09	0.74	0.78	1.00	0.93
F3	0.08	0.98	0.98	0.99	0.98
M3	0.07	0.96	0.94	0.90	0.94
F6	0.06	1.00	1.00	0.99	1.00
M4	0.03	0.97	0.95	0.99	0.97
F5	0.01	0.83	0.65	0.65	0.72
Average		0.97	0.96	0.98	0.97

Table 4. Inter-observer reliability, given as Pearson Correlation Co-efficients, for food type data collected during 1-minute instantaneous recording.

Food type	Proportion	Observer pair 1	Observer pair 2	Observer pair 3	Average
Leaves	0.27	0.99	0.99	0.98	0.99
Fruit	0.23	0.99	0.99	0.97	0.99
Shoot	0.21	0.98	0.96	0.98	0.98
Faeces	0.17	0.99	1.00	0.98	0.99
Ants	0.11				0.00
Roots	0.00				0.00
Average		0.93	0.95	0.91	0.93

Table 5. Comparison of different sample intervals on proportions of behaviours* conducted by the focal gorilla during instantaneous recording, and on inter-observer reliability, given as median values of Pearson correlation coefficients for 3 observers. (*8 most frequently recorded behaviours (of 27 recorded), accounting for over 90% (91.3%) of data recorded during 1-minute interval sampling.)

Behaviour	Proportion			Median Correlation Coefficient		
	1-minute	5-minutes	10-minutes	1-minute	5-minutes	10-minutes
Rest	0.26	0.25	0.26	0.94	0.82	0.62
Eat	0.18	0.16	0.11	0.98	0.93	0.71
Out-of-sight - rest/forage	0.14	0.14	0.13	0.99	0.99	0.99
Play	0.12	0.11	0.12	0.94	0.95	0.86
Sleep	0.10	0.11	0.13	0.99	0.99	1.00
Observe	0.06	0.05	0.06	0.87	0.38	0.53
Move	0.04	0.04	0.04	0.22	0.32	-0.22
Self-play	0.02	0.02	0.03	0.99	0.88	0.65
Median				0.96	0.91	0.68

Table 6. Inter-observer reliability, given as Pearson Correlation Co-efficients, for behaviours* conducted by the focal gorilla during continuous recording. (*8 most frequently recorded behaviours (of 18 recorded, excluding non-target behaviours), accounting for over 90% (92.7%) of data recorded).

	Mean frequency	Observer pair 1	Observer pair 2	Observer pair 3	Median
Chest-beat	3.33	0.95	0.96	0.94	0.95
Charge	1.26	0.98	0.80	0.86	0.86
Tree-beat	0.56	0.67	0.86	0.62	0.67
Hand-clap	0.48	0.81	1.00	1.00	1.00
Hit	0.39	0.40			
Posture	0.33				
Investigate	0.26	0.65	0.65	1.00	0.65
Smash vegetation	0.24	0.58			
Grab	0.22	0.88			
Median		0.74	0.86	0.94	0.86

Table 7. Inter-observer reliability, given as Pearson Correlation Co-efficients, for behaviours* conducted by non-focal gorilla during continuous recording. (*11 most frequently recorded behaviours (of 20 recorded, excluding non-target behaviours), accounting for over 90% (91.9%) of data recorded).

	Mean frequency	Observer pair 1	Observer pair 2	Observer pair 3	Median
Chest-beat	2.35	0.77	0.95	0.77	0.77
Charge	1.57	0.72	0.78	0.68	0.72
Cough	0.52	0.00	0.34	1.00	0.34
Hit	0.48	0.68			
Attack	0.39	0.76	1.00	0.65	0.76
Intervene	0.37	1.00	1.00	1.00	1.00
Posture	0.35				
Tree-beat	0.33	0.80	1.00	0.80	0.80
Scream	0.22	0.94	0.65	0.34	0.65
Chase	0.15	1.00			
Grab	0.15	0.66			
Median		0.77	0.95	0.77	0.76