More on the meanderings of mangabeys: how to test whether bounded walks are random

J. M. C. HUTCHINSON*

School of Mathematics, University of Bristol, Bristol BS8 1TW, UK

Summary

1. Barrett & Lowen (1998) and Waser (1976) attempted to explain the net monthly and yearly displacements of Grey-Cheeked Mangabeys using observed short-term step lengths and assuming a random walk, with and without boundaries. This paper reanalyses their data.

2. Analytic approaches require the root-mean-square step length, not the mean. However, a more flexible approach to making and testing predictions is Monte-Carlo simulation. With a random walk long-term displacements have a large variance, so a single observation is unlikely to disprove this null hypothesis.

3. Restricting movement to a square lattice is a reasonable approximation even when rectangular boundaries are incorporated. Describing the boundary configuration accurately is more important.

4. The observed non-uniformity in turning angles should have been incorporated as it has a large effect on predicted net displacements, unless the arena is tightly constricted. Randomness of movement within a day can be distinguished from that between days. For Waser's population it makes sense to predict long-term displacements using only long-distance daily displacements.

5. There are better approaches to establish both whether boundaries exist and whether movements follow a random walk.

Key-words: Cercocebus albigena, home range, Monte-Carlo simulation, random walk, territoriality

Functional Ecology (2000) **14**, 267–271

Introduction

Recently Barrett & Lowen (1998) reapplied some techniques earlier used by Waser (1976). Both papers tested whether Grey-Cheeked Mangabeys, *Cercocebus albigena* Lydeker, took a random walk through Kibale Forest in Uganda. Waser (1976) recorded a group's position every half-hour, and Barrett & Lowen (1998) recorded group position after every movement over 10 m. They averaged the step lengths between successive positions and used this average to predict monthly and yearly net displacements (i.e. the straight-line distance from start to finish, henceforth written as *r*). Movements were assumed to follow an unbiased and uncorrelated random walk, defined as a sequence of straight-line steps with directions independent of position, step length and the directions of previous steps. In both studies observed monthly and yearly values of *r* were less than predicted.

Barrett & Lowen (1998) tried to account for the disagreement, using both their own data and Waser's. They still modelled movement as a random walk, but the walk was now constrained by two parallel boundaries corresponding to areas of unsuitable habitat. Later they added two boundaries perpendicular to the first, corresponding to the home ranges of neighbouring groups. They claimed that only with all four boundaries were observations and predictions not significantly different, and they argued from this that neighbouring groups indeed formed barriers.

In this paper I identify some methodological weaknesses in this part of Barrett & Lowen's (1998) analysis. My primary aim is not to challenge their biological conclusions, but to improve future testing for deviations from a random walk. Random walks, and the diffusion equations that approximate them, have been used in biology to explain the behaviour of both whole organisms and molecules, and they are incorporated in models of foraging, population dynamics, disease transmission and gene flow (e.g. Okubo 1980). Often a random walk is the null hypothesis, whose rejection would imply more interesting phenomena (e.g. a taxis). So it is important to test correctly whether a random walk adequately describes behaviour.

The formula for root-mean-square net displacement

Equation 1 gives the root-mean-square (RMS) net displacement after an unbounded random walk of *N* steps, each of length *L* (Rayleigh 1880):

*Present address: School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK.

RMS of $r = L\sqrt{N}$. eqn 1

Barrett & Lowen (1998) printed this incorrectly but used the correct version. The formula holds in all dimensions, whether movement is allowed in any direction or restricted to a square lattice.

The same formula holds if step length varies, whatever its distribution, but only if L is redefined as the RMS step length (e.g. Hughes 1995, p. 75). Barrett & Lowen (1998) put $L = 82$ m, but this was the mean step length. I estimate from Barrett's (1995) histogram of step lengths that the RMS step length was about 98 m, making Barrett & Lowen's predictions 16% too low. The 60·85-m step length used by Waser (1974, 1976) is also cited as a mean; however, a histogram of half-hourly movements published later by Waser (1984a) has an RMS of about 61 m, so he may have used the RMS (P. M. Waser, personal communication). Because of this uncertainty about the distribution of step lengths, my predictions for Waser's mangabeys should be regarded as only illustrative.

Barrett (1995) criticized Waser (1976) for basing his calculations on half-hourly displacements rather than the length of distinct movements. If each half-hourly displacement was itself the product of a smaller-scale random walk, Waser's approach is fine. But suppose that independent straight-line movements typically lasted an hour; using half-hourly displacements would then divide the true L by 2 and multiply the true N by 2, leading to an overestimation of r by a factor of $\sqrt{2}$ (equation 1). This would be a special case of the directions of consecutive steps being correlated (see below); Barrett's approach of recording position when the group pauses faces the same problem if movement tends to continue in the same direction after a pause.

Methods of statistical testing

Barrett & Lowen (1998) did not test statistically whether their observations fitted the predictions of their unbounded model, and with their bounded simulation models they inappropriately used Normal-deviate values.

The distribution of r for large N has the following probability density function (Rayleigh 1880):

$$p(r) = \frac{2r}{N} e^{-r^2/N}.$$ eqn 2

This distribution is skewed and that of r^2 even more so (63% of r are below the RMS value). Consequently a t-test is only appropriate if testing the mean of several observations of r, preferably using a transformation. With fewer observations significance levels should instead be calculated analytically from equation 2. The cumulative density function of r is:

$$\text{Probability } (r < R) = 1 - e^{-R^2/N}.$$ eqn 3

We can readily calculate from this the values of r below which 2·5% and 97·5% of the population should lie (the two-tailed 95% confidence limits). This test is algebraically equivalent to looking up observed values of $2r^2/N$ in standard χ^2 tables using df = 2 (these give one-tailed probabilities). If m independent random walks have been observed, look up

$$\sum_{i=1}^{i=m} (2r_i^2/N_i)$$ in χ^2 tables with df = $2m$. Note that equations

2 and 3 hold only if N is large, or if the distribution of step lengths itself follows equation 2 (often a fair approximation – Waser 1984a).

Alternatively one can Monte-Carlo simulate a large number of random walks, and reject the null hypothesis if the observation falls within the bottom or top 2·5% of the generated values of r. Manly (1997) suggested a minimum of 1000 simulations to establish a 5% significance level, and 5000 for the 1% level. Simulation is the method to use for small N or with an amended null hypothesis (e.g. boundaries introduced). If several values of r have been observed, the obvious procedure is to calculate the mean of these m observations, and amend the simulation to generate many analogous means of m values of r, with each value of r generated by an independent random walk. However, with Waser's (1976) data one month's end point was the next's starting point, and Barrett & Lowen (1998) also measured monthly r over consecutive months. When the model includes boundaries, this non-independence affects the variance of the mean of the m months, and thus should be explicitly incorporated into the simulations. Also, in some cases considered below, it made a difference that, after choosing a random starting point, I simulated the walk for a period before starting to measure r. (Depending on how behaviour at the boundaries is modelled, a randomly walking animal need not be equally likely to occur everywhere within the boundaries.) Another advantage of simulation is that one can test predictions of the mean r as well as of the RMS. A mean is closer to most observed values of r than the RMS and both mangabey studies quoted only the mean.

I repeated Barrett & Lowen's (1998) analyses using simulation to establish the significance level (Table 1). The two-tailed P-values are calculated as twice the lesser of the one-tailed values (Manly 1997, p. 72), and based on 10^5 simulations. Although the observed yearly r in Waser's study is much lower than that expected from an unbounded random walk (1870 m vs 5062 m), the difference is not nearly significant. This reflects the skewed distribution of r and its large variance. However, the chances of obtaining a significant result are much higher with monthly r, because this was measured for several months, allowing calculation of a mean, with a lower variance than single observations.

It would be concluded from Table 1 that partial

Table 1. Net displacements (in metres) as observed in the two studies, and the expected values assuming a random walk (RW), with and without boundaries. Boundaries and the number of steps follow Barrett & Lowen (1998). Step-length distributions are taken from Barrett (1995) and Waser (1984a). One month's finishing point is the next's starting point. If a step would cross a boundary, a different random direction is tried instead, with the same step length. Movement is not restricted to a lattice. Two-tailed significance levels (in parentheses) compare observed with predicted values

	Waser		Barrett	
	Monthly (mean of 11)	Yearly	Monthly (mean of 7)	Yearly
Observed	915	1870	714	620
Unbounded RW	1455 (1·0%)	5062 (20%)	1923 (< 0·1%)	6660 (1·4%)
Partially bounded RW	1044 (54%)	3325 (69%)	1520 (0·4%)	4429 (7·7%)
Fully bounded RW	884 (82%)	1399 (57%)	791 (70%)	809 (77%)

boundaries are sufficient to explain Waser's observations, but that full boundaries are still necessary to explain Barrett's observations of monthly r.

Is a square lattice adequate?

Barrett & Lowen (1998) conducted their random-walk simulations on a square lattice. When N is large and the random walk unbounded, the distribution of r is the same for a square lattice as for the continuous case (Rayleigh 1880). However, for a bounded random walk this need not be true. To see why, consider the mangabey group positioned on a boundary lying along one 'street' of the lattice. I model behaviour at a boundary by picking another direction at random if a step would cross the boundary. Then with the square lattice there is a two-thirds chance that the group remains on the boundary, whereas in the continuous case the group will move away, on average by $2/\pi$ of a step. The tendency to stick to boundaries usually increases r (random points within a rectangle are on average nearer than points on its periphery).

In the case of the fully bounded random walk for Barrett's field site (dimensions 2150 m × 1050 m), the predicted mean monthly r was 791 m (not significantly different from the observed value of 714 m: $P = 70\%$). The prediction is slightly increased (806 m) by having a constant step length of 98·09 m (the RMS value). If instead we model a square lattice with this step length, the predicted mean monthly r is a little larger still (818 m; now $P = 59\%$). One can specify other rules for behaviour at boundaries. If boundaries act like idealized billiard cushions, restricting the walk to a lattice has an opposite effect on mean monthly r (798 m *vs* 828 m). But a square lattice seems not to be too misleading with this boundary configuration.

How much does boundary configuration matter?

Barrett & Lowen (1998) modelled the boundaries as straight lines lying parallel and perpendicular – a reasonable first step. However, details of the boundary configuration can matter. An extreme case is a dumb-bell shape in which a narrow corridor connects two blocks of habitat. A random walk is unlikely to enter this corridor, so r tends to be smaller than within a single circle of habitat equal in total area. Larger values of r are expected if the area is arranged as a narrow strip (then the distribution of r approaches that for an unbounded random walk).

Simulation models can readily incorporate real boundary configurations. Barrett (1995) displayed the range of her mangabey group over a year. Since the number of quadrats visited had almost reached an asymptote, I have based the boundary configuration on this home range (smoothing the outlines slightly). It is somewhat dumbbell shaped (350 m wide at its narrowest, 1300 m and 950 m wide at the ends). Predicted mean monthly and yearly r are now 710 m and 748 m. For comparison I simulated walks within a rectangle of the same length and area (1820 m × 750 m). Unexpectedly, predicted mean monthly and yearly r are then smaller (651 m and 654 m). The probable reason is that with the rectangle the simulation's starting point is more likely to lie centrally than with the dumbbell, and such starting points preclude larger values of r.

Analysing randomness at different temporal scales

Barrett & Lowen (1998) and Waser (1976) used observed step lengths of a few tens of metres to predict both monthly and yearly values of r. Another possibility would be to predict yearly r from the observed monthly r. But a month is rather an arbitrary duration, and it seems more meaningful to use step lengths or half-hourly displacements to predict daily r, and use observed daily r to predict monthly and yearly r. It is plausible that movements within a day do not fit a random walk, but that monthly r is well predicted by a random walk based on observed daily r (e.g. Jones *et al.* 1980).

Waser (1977a) published a histogram of 108 values of daily r. Their mean is about 484 m. The mean daily r predicted on the basis of an unbounded random walk of 24 steps with lengths distributed as in Waser (1984a) is 265 m. So movement within a day appears

more directional than expected ($P < 0.1\%$). In contrast, given that daily r follows the distribution in Waser's histogram, an unbounded random walk predicts a mean monthly r of 2769 m. This significantly exceeds the observed mean of 915 m ($P < 0.1\%$), and is also much higher than the 1455 m predicted from half-hourly displacements.

However, Waser's (1977a) description of the biology suggests that the null hypothesis in all the above analyses misrepresents the time scale of the random process. The mangabeys fed mostly on fruits of large trees that were well dispersed and fruited out of synchrony (Waser 1974). A tree could fruit for 10 days. When a mangabey group found such a tree, Waser (1977a) had the impression that they returned to it daily, and during the rest of the day rather systematically covered the surrounding area. When the tree stopped fruiting they set off further afield to find another large tree in fruit. Consequently daily r was distributed as two non-overlapping peaks either side of 1000 m (Fig. 9 of Waser 1977a). A more realistic random-walk model to predict yearly movements might incorporate only long-distance movements between these large fruiting trees, ignoring movements under 1000 m, which we suppose are mostly to-and-fro movements around a fruiting tree. Long-distance movements occurred on 34 days per year, and Waser's histogram gives the distribution of this subset of daily r. Predicted mean yearly r is now 6432 m, which is much larger than the group's home range, although still not significantly different from the observed value of 1870 m ($P = 13\%$). (To predict monthly r it would be important that the simulation incorporate variation in the number of days of long-distance movements per month – difficult, since such days did not occur at random.)

Barrett's (1995) mangabeys did not seem to return to the same major food resource over successive days.

Correlated random walks

A tendency to return to one tree is one reason that step direction may violate the assumption of being fully random. Another likely reason is avoiding an area just visited. Indeed Waser (1984b) and Barrett (1995) found a strong tendency for both mangabey groups not to track back in the short term. Given that, the analysis should model movement as a correlated random walk, where directions of successive steps are correlated. A simple elaboration of equation 2 gives the RMS r in such cases (Kareiva & Shigesada 1983), but a computer algorithm is required to calculate its variance (McCulloch & Cain 1989). Instead I used simulation to test for significance. I used the observed turning angles from Barrett (1995), assuming no correlation with step length and a symmetric distribution to left and right. For an unbounded walk, the predicted mean monthly and yearly r are then 2973 m and 10315 m. These considerably exceed the predictions for an uncorrelated random walk (1923 m and 6660 m).

If we add boundaries, behaviour at the boundaries will cause the turning angles in the simulation to diverge from the specified distribution. But if we ignore this problem, the predicted mean monthly and yearly r in the fully bounded case are 841 m and 843 m for the correlated random walk, not much greater than the 791 m and 809 m for the uncorrelated random walk. In fact the gross path length in a year is so great relative to this arena that all sorts of random movement rules yield similar values of r, because the end position is almost independent of the starting position. Two points chosen randomly within the arena lie on average 860 m apart.

Centrally biased random walks

The hard boundaries used by Barrett & Lowen (1998) are appropriate for the sharp habitat transitions described by their partially bounded model. However, the social boundaries with neighbouring groups seem to be softer, since group ranges overlap extensively (Waser 1976). Also, Barrett (1995) describes a core area that her mangabeys occupied more frequently. Thus a more realistic model might be a centrally biased random walk in which the probability of moving towards the centre increases with distance from the centre (for examples see Chapter 8 of Okubo 1980). The problem is deciding on an appropriate function relating central bias to position.

Conclusions

The important points to emerge are that to predict long-term net displacements from short-term movements one should: (i) either simulate with the observed distribution of step lengths or use their RMS; (ii) describe boundaries accurately; and (iii) either choose a time scale so that consecutive movements are uncorrelated in direction, or incorporate the observed turning angles. Approximating the walk on a square lattice was surprisingly accurate. Differences between observed and expected values of r are unlikely to be significant unless several observed values are averaged.

More generally this example demonstrates the danger of taking simple analytic models 'off the peg' without appreciating their underlying assumptions, which may severely restrict their applicability. Numerical simulation allows random-walk models to be applied more flexibly, but also shows that predictions can be sensitive to the assumptions. Accurate predictions modelled closely on the biology would require much more information, perhaps more than is feasible to collect. But the purpose of models is not always to make accurate predictions, and here random walks have provided an example to demonstrate the importance of boundaries in restricting displacements.

It is unsurprising that when the model incorporates observed boundaries the predicted net displacements fit the observations better. As an argument for the

existence of social boundaries the procedure is somewhat circular: the same sequence of movements that gave the displacements also suggested the boundary locations, so they cannot be too incompatible. As an argument for the mangabeys taking a random walk it is weak: if the boundaries are restricted enough relative to the gross path length, all sorts of movement rules will result in similar net displacements. There are better, more direct, ways to test whether trajectories are true random walks, for instance by examining the autocorrelation of turning angles. Whether boundaries exist is better investigated by focusing on behaviour near the proposed boundary. For instance, Waser (1977b) went on to perform experimental playbacks of mangabey calls; mangabeys avoided the speaker, but only if under 200 m away, and avoidance was no greater near the edge of their home range.

Acknowledgements

References

Barrett, L. (1995) *Foraging strategies, range use and territoriality of grey-cheeked mangabeys (*Cercocebus albigena*) in Kibale Forest, Western Uganda*. PhD Thesis, University of London.

Barrett, L. & Lowen, C.B. (1998) Random walks and the gas model: spacing behaviour of Grey-Cheeked Mangabeys. *Functional Ecology* **12**, 857–865.

Hughes, B.D. (1995) *Random Walks and Random Environments, Vol. 1: Random Walks*. Oxford University Press. Oxford.

Jones, R.E., Gilbert, N., Guppy, M. & Nealis, V. (1980) Long-distance movement of *Pieris rapae*. *Journal of Animal Ecology* **49**, 629–642.

Kareiva, P.M. & Shigesada, N. (1983) Analyzing insect movement as a correlated random walk. *Oecologia* **56**, 234–238.

McCulloch, C.E. & Cain, M.L. (1989) Analyzing discrete movement data as a correlated random walk. *Ecology* **70**, 383–388.

Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology* 2nd edn. Chapman & Hall. London.

Okubo, A. (1980) *Diffusion and Ecological Problems: Mathematical Models*. Springer-Verlag. Berlin.

Rayleigh (1880) On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philosophical Magazine, Series 5* **10**, 73–78.

Waser, P.M. (1974) *Intergroup interaction in a forest monkey: the mangabey* Cercocebus albigena. PhD Thesis, Rockefeller University, New York.

Waser, P.M. (1976) *Cercocebus albigena*: site attachment, avoidance, and intergroup spacing. *American Naturalist* **110**, 911–935.

Waser, P. (1977a) Feeding, ranging and group size in the mangabey *Cercocebus albigena*. *Primate Ecology: Studies of Feeding and Ranging Behaviour in Lemurs, Monkeys and Apes* (ed. T. H. Clutton-Brock), pp. 183–222. Academic Press, London.

Waser, P. (1977b) Individual recognition, intragroup cohesion and intergroup spacing: evidence from sound playbacks to forest monkeys. *Behaviour* **60**, 28–74.

Waser, P.M. (1984a) 'Chance' and mixed-species associations. *Behavioral Ecology and Sociobiology* **15**, 197–202.

Waser, P.M. (1984b) Ecological differences and behavioral contrasts between two mangabey species. *Adaptations for Foraging in Nonhuman Primates* (eds P. S. Rodman & J. G. H. Cant), pp. 195–216. Columbia University Press, New York.