A PROBLEM OF BIAS IN SCAN SAMPLING

Kristen Hawkes

Department of Anthropology, University of Utah, Salt Lake City, UT 84112

Hillard Kaplan

Department of Anthropology, University of New Mexico, Albuquerque, NM 87131

Kim Hill and Magdalena Hurtado

Department of Anthropology, Emory University, Atlanta, GA 30322

Instantaneous scan sampling is intended to accumulate data sets large enough to allow quantitative assessments of activity budgets. If all subjects are not continuously visible, group activities may be seriously overrepresented in the sample of observations collected by scans. This paper describes the problem and our solution.

INSTANTANEOUS SCAN SAMPLING, also called spot observation and point sampling, has become a standard technique in studies of animal behavior (Altmann 1974; Dunbar 1976). Anthropologists and cross-cultural psychologists (Blurton Jones 1972) have increasingly adopted it to collect quantifiable records of everyday behavior in small-scale societies (e.g., Munroe and Munroe 1971; Johnson 1973; Nerlove et al. 1974; Draper 1975; Konner 1976; Hames 1979; Gross et al. 1979; Werner et al. 1979; Rogoff 1981; Betzig and Turke 1985; see review in Gross 1984; critique in Borgerhoff Mulder and Caro 1985). But the resulting data set can be compromised by a systematic bias inherent in the technique when it is used where subjects are not continuously visible: spot observations consistently overrepresent group activities.

The technique of spot observation varies in some details among investigators but generally consists of sequences of observations (at randomly scheduled beginning times) of a random or stratified sample of the population of interest. The observer notes the pertinent features of the activity of the subject at the moment he or she is first observed. Beginning times and often the order in which subjects are observed are randomized so that data sets are representative of the entire time period of interest. However, a bias appears in the data if the time interval between each individual spot observation varies widely according to the time required to search out the subject. Certain behaviors (particularly group activities involving a large subset of the sample population) and certain time periods (those in which group activities occur) can be massively overrepresented. We became aware of the bias in observations collected by this technique during recent fieldwork.[1]

Our study population consists of about two hundred Ache Indians currently residing part of the time at a mission-sponsored agricultural settlement, Chupa Pou, established in 1978 in eastern Paraguay. Until 1972 the hunting and gathering northern Ache had no unarmed contact with outsiders, and some of them remained full-time foragers until they settled at Chupa Pou. Here Ache residents have built houses and cultivate fields of corn, manioc, and sugarcane. They also grow some sweet potatoes and bananas and keep a few chickens and pigs. Many residents continue to spend part of their time on foraging trips away from the settlement.

On the basis of K. Hill's friendships and experience from earlier work with the Ache, in 1980 we began to systematically study Ache foraging (Hawkes, Hill, and O'Connell 1982; Hill and Hawkes 1983). We returned in 1981–82 to continue monitoring hunting and gathering (Hill et al. 1984, 1987) and to collect data on patterns of sharing (Kaplan et al. 1984; Kaplan and Hill 1985), time allocation (Hill et al. 1985; Hurtado et al. 1985), the archaeological reflection of Ache life in the forest (Jones 1983), and settlement activities (Hawkes et al. n.d.).

Initially we intended to use instantaneous scan sampling to compile a record of settlement behaviors which would allow us to construct activity budgets for time spent at the colony (ideally for individuals, but at least for sex, age, and other social categories). However the rhythm of settlement life revealed a crippling bias in data collected by this technique. Inevitably the record of observations would overrepresent group activities in a setting, like this one, where people may congregate or they may disperse far outside visual range of each other.

One kind of settlement activity makes this problem particularly clear. Almost daily some men play soccer and/or volleyball before an audience of mixed sexes and ages. If the period during which scanning is scheduled overlaps a game, many subjects will immediately be recorded as playing or watching. In other words, one observation of playing soccer will be multiplied several times very quickly because many of the subjects are found with no time spent searching.

In a very simplified fashion this is illustrated more formally by the following model. Assume a population of eight individuals (1–8), an array of five possible activities (A–E), and a very short day divisible into four time periods (I–IV) during which these activities occur. All the activities except D require the observer to search for a subject, a search which, for simplicity, shall be assumed to take one of the time periods to complete. D is exceptional in that it is a group activity. Once anyone doing D is found, the activity of all other subjects also engaged in D is directly visible. The actual distribution of activities is shown in Table 1. Columns represent subjects, rows are time periods, and the entries in the cells of the matrix are activities. Assume that all days are identical to the one represented in this table.

Suppose that we wish to know the proportion of time subjects spend at various activities. We employ a version of scan sampling, simplified in that subjects are always taken in the same order, and a series of four observation sequences is begun at each of the time periods in straight rotation. Follow the rule that if all subjects have not been observed before the time of interest ends for the day (perhaps it gets dark), the sequence will be resumed at the beginning of the following day and carried to completion.

TABLE 1
Actual Occurrence of Activities

	Subjects							
Time Periods	1	2	3	4	5	6	7	8
I	A	B	C	E	A	B	C	E
II	B	C	E	A	D	D	D	D
III	C	E	A	B	B	C	E	A
IV	E	A	B	C	C	E	A	B

Activity	Frequency	Proportion of Total
A	7	0.22
B	7	0.22
C	7	0.22
D	4	0.12
E	7	0.22

Table 2 shows the observation record which results. Columns again represent subjects, rows here represent beginning time periods, and the entries in the cells of the matrix are the activities observed. For example, the first observation sequence begins at time period I. (Refer to Table 1 for the actual distribution of activities.) Subject 1 is found engaged in activity A. The ethnographer then seeks subject 2; by the time he is found, one time period has elapsed, and 2 is engaged in activity C. The ethnographer then looks for subject 3, who, when found, is doing A. The search for subject 4 follows, and 4 is seen to be doing C. The sequence is continued the following day to complete the rotation of subjects. Five is found engaged in A. Then 6 is sought and found doing D. Now, however, as the ethnographer turns to 7, he need not search. Seven is immediately observed, engaged with 6 in D, as is 8. The next observation sequence begins at time period II, with results as recorded in Table 2. When observations have been initiated at each of the time periods, the data show a marked bias. Activity D is overrepresented more than twofold. All other activities are consequently underrepresented.

To prevent this bias, we adopted a spacing rule which was very expensive of time. This rule adds a start restriction to the stop restriction investigators using spot observation must always adopt. The standard rule is of the form: If subject X is not found within some specified time limit, this fact is recorded, and the ethnographer moves on to the next subject. Our spacing rule took the following form: The observation of subject X must occur within the appropriate ten-minute period (determined by the randomized matching of subjects and

TABLE 2
Observed Occurrence of Activities

Observation Sequences	Beginning Time Periods	Subjects							
		1	2	3	4	5	6	7	8
1st	I	A	C	A	C	A	D	D	D
2nd	II	B	E	B	E	D	D	D	D
3rd	III	C	A	C	A	B	E	C	D
4th	IV	E	B	E	B	C	B	D	D

Activity	Observed Frequency	Proportion of Total
A	5	0.16
B	6	0.19
C	6	0.19
D	10	0.31
E	5	0.16

times for any given observation sequence). If he is not found during this period, record that and move on. If he *is* found, do not begin to observe the next subject until the time assigned to that subject begins. If this rule is applied to the simple situation represented in Table 1, the bias shown in Table 2 disappears. By following the rule, the ethnographer, on finding one subject playing D, may not begin observation of the next subject until the next time period commences, at which time that subject has left D for some other activity.

Unfortunately, this correction does not solve the problem in data sets already collected. However, such data sets can still be used to show relative differences among some subjects within a study population: to the extent that these differences are not systematically associated with differences in participation in group activities, the bias discussed here will not distort results. For example, differences between male and female subjects in the proportion of scans which find them grooming may reflect actual differences. But the result might be spurious if grooming is more or less likely in *large* groups and one sex tends to congregate periodically.

Biases toward some observation contexts may be located in a data set by arraying it to display differential sampling. For the record shown in Table 2, this would mean mapping the entries onto a matrix which distinguishes not only subjects and the observation sequence beginning time periods, but also the time period at which each spot was recorded. Table 3 shows such a display. Here the rows are occurrence or observation time periods. Column clusters are observation sequences, identified by beginning times (these appeared as

TABLE 3
Biased Sampling Displayed

First Observation Sequence

Time Period of Observation	Subjects							
	1	2	3	4	5	6	7	8
I	A	—	—	—	A	—	—	—
II	—	C	—	—	—	D	D	D
III	—	—	A	—	—	—	—	—
IV	—	—	—	C	—	—	—	—

Second Observation Sequence

Time Period of Observation	Subjects							
	1	2	3	4	5	6	7	8
I	—	—	—	E	—	—	—	—
II	B	—	—	—	D	D	D	D
III	—	E	—	—	—	—	—	—
IV	—	—	B	—	—	—	—	—

Third Observation Sequence

Time Period of Observation	Subjects							
	1	2	3	4	5	6	7	8
I	—	—	C	—	—	—	C	—
II	—	—	—	A	—	—	—	D
III	C	—	—	—	B	—	—	—
IV	—	A	—	—	—	E	—	—

Fourth Observation Sequence

Time Period of Observation	Subjects							
	1	2	3	4	5	6	7	8
I	—	B	—	—	—	B	—	—
II	—	—	E	—	—	—	D	D
III	—	—	—	B	—	—	—	—
IV	E	—	—	—	C	—	—	—

Time Period	Number of Observations	Proportion of Total
I	7	0.22
II	14	0.44
III	5	0.16
IV	6	0.19

rows in Table 2). Each cell in this matrix defines a unique time period, observation sequence, and subject. The display shows that time periods have been very differently sampled. Different degrees of confidence are accordingly appropriate, and single observations carry very different weights. Fourteen observations occur within time period II. The first four, those which occur during the first observation sequence, are jointly equivalent to the observations which occur within each of the other time periods during that sequence. They are worth only "one-quarter" each, as the two observations in time period I during that observation sequence are worth "one-half" each. In the second observation sequence, the five observations in time period II are worth only "one-fifth" each.

In a larger data set, the number of observations in time periods which follow group activities would be sharply reduced because the period in which the group activity begins absorbs so many observations. This entails another bias. With the collapse of obvious duplicates, group activities may now be *under*-represented. A time period which follows observations of group activities will be *systematically* unlikely to include spots of those same group activities. Even if these continue through subsequent time periods, the observer will have quickly finished scanning subjects in groups. This means that observations will be recorded in the following time period if, but only if, subjects are doing something else.

Certainly spot observation has many advantages over traditional ethnographic techniques for assessing time allocation and behavioral variation, which have ranged from baldly asserted generalizations, through illustrative anecdotes, to unsystematic tabulations (see discussion in Johnson 1978; Borgerhoff Mulder and Caro 1985). But the bias described here may be a source of significant distortion in time budgets calculated from instantaneous scan data collected in settings where subjects may alternately cluster or disperse out of sight.

NOTE

1. The research on which this report is based was funded by the University of Utah Research Committee, the Leakey Foundation, the National Science Foundation (#BNS-8309834), and the National Institutes of Health (#1 R01 HD16221-01A2).

REFERENCES CITED

Altmann, J., 1974, Observational Study of Behaviour: Sampling Methods. Behaviour 49:227–67.

Betzig, L.L., and P.W. Turke, 1985, Parental Investment by Sex on Ifaluk. Ethology and Sociobiology 7:29–37.

Blurton Jones, N.G., 1972, Ethological Studies of Child Behavior. Cambridge, Eng.: Cambridge University Press.

Borgerhoff Mulder, M., and T.M. Caro, 1985, The Use of Quantitative Observational Techniques in Anthropology. Current Anthropology 26(3):323–35.

Draper, P., 1975, Cultural Pressure on Sex Difference. American Ethnologist 4:600–16.

Dunbar, R.I.M., 1976, Some Aspects of Research Design and Their Implications in the Observational Study of Behavior. Behaviour 58:78–98.

Gross, D.R., 1984, Time Allocation: A Tool for the Study of Cultural Behavior. Annual Review of Anthropology 13:519–58.

Gross, D.R., G. Eiten, N.M. Flowers, F.M. Leoi, M. Lattman Ritter, and D.W. Werner, 1979, Ecology and Acculturation among Native Peoples of Central Brazil. Science 206:1043–50.

Hames, R.B., 1979, Relatedness and Interaction among the Ye'kwana: A Preliminary Analysis. Pp. 238–49 in Evolutionary Biology and Human Social Behavior: An Anthropological Perspective (ed. by N.A. Chagnon and W. Irons). North Scituate, Mass.: Duxbury Press.

Hawkes, K., K. Hill, and J.F. O'Connell, 1982, Why Hunters Gather: Optimal Foraging and the Ache of Eastern Paraguay. American Ethnologist 9:379–98.

Hawkes, K., H. Kaplan, K. Hill, and A.M. Hurtado, n.d., Ache at the Settlement: Contrasts Between Farming and Foraging. Human Ecology. In press.

Hill, K., and K. Hawkes, 1983, Neotropical Hunting among the Ache of Eastern Paraguay. Pp. 139–88 in Adaptations of Native Amazonians (ed. by R. Hames and W. Vickers). New York: Academic Press.

Hill, K., K. Hawkes, A.M. Hurtado, and H. Kaplan, 1984, Seasonal Variance in the Diet of Ache Hunter-Gatherers in Eastern Paraguay. Human Ecology 12:145–80.

Hill, K., H. Kaplan, K. Hawkes, and A.M. Hurtado, 1985, Men's Time Allocation to Subsistence Work among the Ache of Eastern Paraguay. Human Ecology 13(1):29–47.

Hill, K., H. Kaplan, K. Hawkes, and A.M. Hurtado, 1987, Foraging Decisions among Ache Hunter-Gatherers: New Data and Implications for Optimal Foraging Models. Ethology and Sociobiology 8:1–36.

Hurtado, A.M., K. Hawkes, K. Hill, and H. Kaplan, 1985, Female Subsistence Strategies among the Ache of Eastern Paraguay. Human Ecology 13(1):1–28.

Johnson, A., 1973, Time Allocation in a Machiguenga Community. Ethnology 14:301–10.

Johnson, A., 1978, Research Methods in Social Anthropology. London: Edward Arnold.

Jones, K., 1983, Forager Archaeology: The Ache of Eastern Paraguay. Pp. 171–91 in Carnivores, Human Scavengers, and Predators: A Question of Bone Technology (ed. by G.M. Lemoyne and A.S. MacEachern). Calgary: University of Calgary Archaeological Association.

Kaplan, H., and K. Hill, 1985, Food Sharing among Ache Foragers: Tests of Explanatory Hypotheses. Current Anthropology 26(2):131–33.

Kaplan, H., K. Hill, K. Hawkes, and A.M. Hurtado, 1984, Food Sharing among Ache Hunter-Gatherers of Eastern Paraguay. Current Anthropology 25(1):113–15.

Konner, M.J., 1976, Maternal Care, Infant Behavior and Development among the Kung. Pp. 218–45 in Kalahari Hunter-Gatherers: Studies of the Kung San and Their Neighbors (ed. by R.B. Lee and I. De Vore). Cambridge, Mass.: Harvard University Press.

Munroe, R.H., and R.L. Munroe, 1971, Household Density and Infant Care in an East African Society. Journal of Social Psychology 85:9–13.

Nerlove, S.B., J.M. Roberts, R.E. Klein, C. Yarbrough, and J.P. Habicht, 1974, Natural Indicators of Cognitive Development: An Observational Study of Rural Guatemalan Children. Ethos 2:265–95.

Rogoff, B., 1981, Adults and Peers as Agents of Socialization: A Highland Guatemalan Profile. Ethos 9:18–36.

Werner, D., N.M. Flowers, M. Lattman Ritter, and D.R. Gross, 1979, Subsistence Productivity and Hunting Effort in Native South America. Human Ecology 7:303–15.