Studies in Avian Biology No. 6:366–371, 1981.

REDUCING BIRD COUNT VARIABILITY BY TRAINING OBSERVERS

CAMERON B. KEPLER[1] AND J. MICHAEL SCOTT[2]

ABSTRACT.—During avian surveys, an important variable that affects our ability to determine such basic bird population parameters as species diversity, distribution, density, and population size is the competency of the observers. Skills such as visual and audio acuity, familiarity with the birds being counted, motivation, and willingness to make identifications must all be seriously considered when designing any avian survey. The importance of training observers and ways to do so are discussed in detail.

During a 5-year survey of the forest birds of Hawaii, we developed a 3-week training program that attempts to minimize the variability inherent in a team that has included more than 20 observers.

Our training program included screening of applicants for the physical, psychological, and academic skills mentioned above. We then provide successful applicants with visual (study skins, field guides, slides, etc.) and auditory aids (tape recordings) of species anticipated during the survey. Time is then spent camping under field conditions to provide maximum exposure to the birds. Two to four new team members train with each experienced observer. Problem species are dealt with on an individual basis. Simultaneous counts are used to determine progress. As identification skills increase, the number of species is increased and distance estimates are added.

Among the many variables that stand between the avian biologist and an understanding of such basic population parameters as species diversity, distribution, density, and population size is the competence of the bird observer. It is clear that all birders are not equal in experience with nor ability to distinguish between or correctly identify all species that may occur in a study area (Enemar 1962, J. T. Emlen 1971, Berthold 1976, Carney and Petrides 1957). Visual acuity, color sensitivity, peripheral vision, and hearing acuity are all important physical characters that vary among observers. Such psychological factors as concentration, motivation, attention span, alertness, endurance, ability to remember nuances of color, shape, or sound, and willingness to make identifications can all critically affect census results. Even the same observer can vary from day to day, or hour to hour, in any or even all of these factors. Berthold (1976), citing numerous European studies, notes that variations in estimates of numbers among observers amounting to 50% or more are common for many, or in some cases even all, species in a study area. The complexity of this problem has led some biologists to forsake avian censuses altogether (Berthold 1976), and others to suggest that uniformity and comparability of results can be achieved only by using the same set of observers in different areas, seasons, or years.

The need for comparability among observers is thus great. That this can be achieved through actual training has only been alluded to in the past (Svensson 1977a, Kimball 1949), while the use of "experienced" observers has been more frequently suggested as the solution to observer problems (Carney and Petrides 1957). However, even experienced birders vary considerably in their abilities, and we suggest that their lack of comparability is still a serious problem. Without accurate identification of species all other efforts to improve the reliability of bird counts are fruitless.

During the course of the five year Hawaiian Forest Bird Survey we became increasingly aware of the importance of carefully selecting and training observers. We have attempted to reduce their inherent diversity by screening all applicants, then involving the selected team members in an intensive training program. We have found that training is a critically important component of our overall experimental design, and describe procedures in this paper that reduce observer variability. The methods we describe have evolved over a five year period.

THE TRAINING PROGRAM

SELECTING OBSERVERS

Observers vary consistently in such physical characteristics as peripheral vision, color blindness, and visual and aural acuity. All applicants to survey programs who are known to have been exposed to hearing stresses (scuba divers, hunters, ex-military men, etc.) should be carefully screened before acceptance. Because hearing loss is in part age-dependent, young observers are generally better than older ones. All potential team members should have hearing tests (audiograms), from 1 to 8 KHz, and those with

[1] U.S. Fish & Wildlife Service, Patuxent Wildlife Research Center, Maui Field Station, 248 Kaweo Place, Kula, Hawaii 96790.

[2] U.S. Fish & Wildlife Service, Patuxent Wildlife Research Center, Mauna Loa Field Station, P.O. Box 44, Hawaii National Park, Hawaii 96718.

serious hearing loss (20 db or greater?) within the range of bird vocalizations used for non-counting tasks or rejected. Years of experience with species may compensate for partial hearing loss and this should be considered when evaluating observers (Emlen and DeJong 1981). Faulty vision, if corrected, is not a problem. However, it may be appropriate to test for tunnel vision.

Psychological factors of known importance, such as motivation, attention span, and willingness to make identifications, should also be considered in the initial screening. For survey work in remote areas requiring camping, factors such as camping and hiking experience, and willingness to remain away from family or the comforts of civilization, should be weighed heavily in the selection process. While birding experience is a key selective factor, we have found that familiarity with the species occurring in the area is not, if the avifauna is simple. Inexperienced observers can be trained, and they learn rapidly.

Laboratory Training

All new observers should begin training indoors. This ordinarily involves an initial full day, followed by short review sessions during the field training period. Trainees should receive field guides and lists of anticipated species. Handouts that identify problem species or confusing groups of species can be provided. Other essential written materials include relevant publications, details of survey methodology, and background natural history information on the area's avifauna.

The heart of the laboratory session should involve an inspection of study skins of all anticipated species, including various morphs, age, and sex classes. Observers can then work with the skins in conjunction with field guides and other written materials. During this session informal spot quizzes and comparisons between similar species can be made. Slides detailing birds, forest types, terrain, and former survey activities are extremely useful, and provide a forum for informal tests that sharpen the participants' skills, especially after their initial exposure to the skins, and some field experience. This material needs to be available throughout the training period, and as observers learn more in the field, they can return to review it.

Tape recordings of songs and call notes should also be provided. Species can be arranged phylogenetically as well as by similarity of sound. When possible, training tapes should be of birds actually recorded in the study area to avoid confusing dialects. Observers need to listen to these tapes throughout the training session, and during the field season when questions arise.

Field Training

From the lab, trainees can be taken into the field and provided with maximum exposure to the birds by camping in the study area. Selection of a site with a small number of species will allow trainees to focus their attention. The first day divide trainees into small groups (two to five) led by an experienced observer. The leader can point out and identify as many birds as possible: "That's an Apapane, that's an . . . , notice the wing bars, etc." The initial objective is to maximize the trainee's exposure to a single set of species. This experience can be augmented at camp by listening to tapes and checking field guides and other identification aids. Trainees should be encouraged to ask questions. A frank statement that even the most experienced observer can't identify every bird is needed. This guided training should be continued as long as necessary, for it offers an excellent opportunity for the new observer to ask "what's that? Why? How does it differ from . . . ?"

Simultaneous counts

Trainees learn a reasonable number of calls and songs in the first day or two. After this the most important training exercise—simultaneous 5-minute counts—should be introduced. An experienced observer and 2 to 5 trainees simultaneously, and independently, record all species seen or heard from one spot (station). At the end of each count the different species lists are compared. These verbal exchanges strongly reinforce the data, and problem species are quickly identified whether they are missed or misidentified. We have consistently found that some common and conspicuous species are missed, even though trainees can readily identify them when they are pointed out. We call these birds "window species," for observers listen right through them without detecting them. Most observers have their own unique constellation of window species, and usually only realize this when they repeatedly fail to detect species that other observers consistently record. The window species are thereafter stressed on an individual basis, and trainees enthusiastically help each other.

Simultaneous counts are best conducted at the start of each day's work. Trainees can then identify species that are giving them difficulty, and work on them later that day. The whole process is simply one of focusing on an increasingly smaller number of problem species. The importance of mixing experienced and inexperienced observers can not be overstressed. Experienced observers provide instant feedback as to the identity of a bird, especially during camping, when everyone wakes up, eats, and drinks, to bird vocalizations. This provides maximum exposure at a time when there is no substitute for field experience.

Most bird survey teams are composed of a mix of new and old observers. Observers from previous seasons need some training, as skills erode with time, but normally do not need an entire training course. They can be brought into the training program at any time, or used as additional teachers. Trainees progress at different rates. Quick learners can be used as part-time trainers to increase their own accuracy and confidence. This is particularly valuable in afternoon sessions, where observers work "one-on-one" to point out to each other the window species discovered during the morning's simultaneous counts.

Simultaneous counts should be used continuously throughout the training period to identify problem species, either those with which trainees lack experience, are overlooking, or are ignoring, for whatever reasons. They are also an excellent measure of progress, provide important motivation, and prepare group members for the actual survey.

When trainees feel comfortable identifying and recording species, they should be asked to record the number of individuals detected during count periods. At this time tell trainees how to search an area around the station. Make sure that observers check directly overhead, rotate 360° if variable circular counts are used (Reynolds et al. 1980), and vary the distance being scanned. Searching at fixed distances is to be avoided as this biases density estimates.

It is important to shift training sites to add new species, terrain, and habitat types. This also improves training efficiency by reducing predictability and monotony. Trainees are usually ready to move to new areas every three or four days.

Distance estimation

Estimating or measuring the distance to detected birds is an integral part of most attempts to determine their numbers. It is critical that these measurements be as accurate as possible. Indeed, Burnham et al. (1980) suggest that "tape measure" precision is required. The use of range finders and flagging at known distances from the observer helps when birds can be seen. However, most observations are of birds heard and not seen (81% of all detections during the 1980 Maui Forest Bird Survey). This requires that distances be estimated. Under ideal conditions, practice can lead to ±10% accuracy (Scott et al. 1981b).

When trainees competently record species and individuals, it is time to introduce them to distance estimation. Working in groups of two to five, they should independently estimate distances to trees, rocks, and other clearly seen fixed objects, then measure the actual distance using tape measures or ranges finders. This exercise must be repeated until trainees are consistently within 10% to 15% of the measured distance. They can then begin estimating the distance to birds seen and heard, then to birds heard only. One observer identifies an unseen vocalizing bird, indicating its direction to the other trainees. They independently estimate its distance. One member of the group then locates it, moves to a point directly under it, and measures the linear distance back to the other observers. Care must be taken to make certain that the bird has not moved before its distance is measured, and that it is the individual originally identified. This exercise is combined with the simultaneous counts and identification work with selected species. Playbacks from hidden tape recorders might also be used to estimate distances (Emlen and DeJong 1981). Distances should be estimated with as many different calls, songs, and species as possible. After trainees achieve an average ±10% accuracy, distance estimation becomes a part of the simultaneous counts.

Beginning the survey

When trainees know the local birds and have mastered distance estimation, they should begin counting under actual survey conditions. Have two observers simultaneously census the same stations or strips to check their progress. When this exercise is concluded, check the field forms and ask the observers if they feel competent to collect real data—their own confidence is important. Quick learners may begin the actual survey while slower trainees continue parts of the training program. Any necessary additional training should be alternated with practice surveys until all trainees are ready to begin the actual counts. In Hawaii, the entire training program takes 12 to 15 days, with experienced observers from former years joining the training session on day 6.

During the actual survey, observers should refer to field guides, tapes, and other aids in order to remain sharp, and to sort out occasional unidentified birds. In the 1980 Hawaii Forest Bird Survey (on Maui) pairs simultaneously censused together (Scott and Ramsey 1981b). For the first four weeks, each pair consisted of an experienced and a new observer, and their interactions after each count provided instant feedback on uncertain species. Thus, in a very real sense, the initial training period extends throughout the field season.

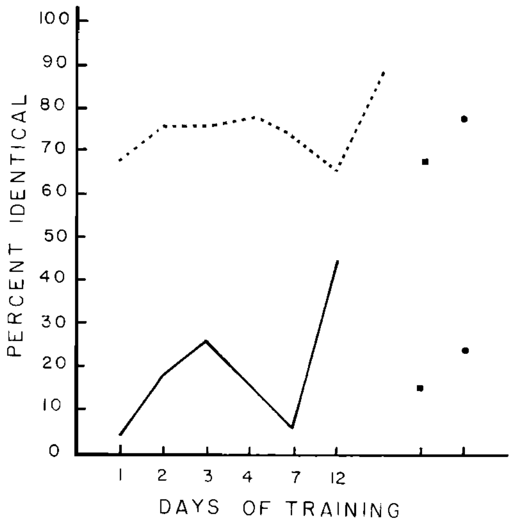

FIGURE 1. Measures of similarity between paired observers. The solid line indicates percent identical species lists, the dotted line the percentage of species simultaneously recorded on the same station counts, Maui Forest Bird Survey training session. The symbols represent the same measures for randomly selected station counts for all Maui (squares) and for ohia forest only (circles) during the actual survey.

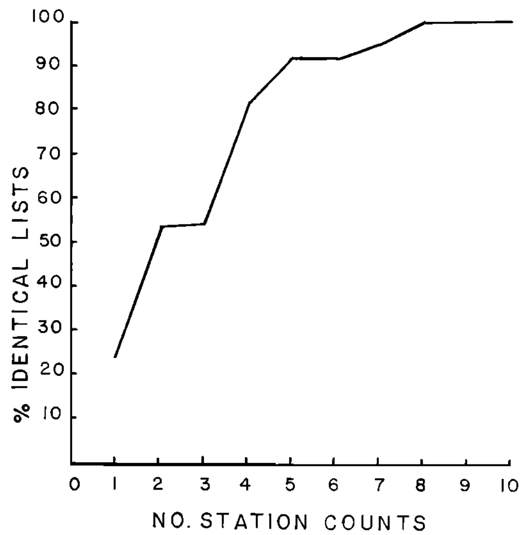

FIGURE 2. The frequency of identical species lists between paired observers as a function of the number of stations sampled, Maui Forest Bird Survey, 1980.

MEASURING OBSERVER VARIABILITY

The following data on bird count variability were taken from simultaneous 5-minute counts made during the 1980 Maui Forest Bird Survey training session, and from actual paired station counts during the survey. The training session counts were simultaneous in time and space, with all observers standing within about 3 m of a central point. They recorded all birds heard and seen, and, on day 7, added distance estimates to each observation. During the survey, observers stood 18.3 m apart, one 9.2 m upslope, the other 9.2 m downslope, from a central station, and this separation contributed importantly to the differences between them. Count periods were eight minutes, and distances to each detected bird were estimated.

Congruity of Species Lists

The simplest measure of similarity between observers is a comparison of species lists. The results can be unexpected. Robbins and Stallcup (1981) matched paired observers on 100 3-minute Breeding Bird Survey stations in Maryland and found no identical lists. Only when 150 stations were compared did they find any (2) that were the same.

The Maui Forest Bird Survey training program began with very low comparability (3%) on the first day. These observers were not naive on day 1. Two of them had participated in the Hawaii Forest Bird Survey for at least two years, on other islands and all others had studied skins, field guides, and tapes. They improved rapidly with training (Fig. 1), reaching a peak (46%) on the last day. The dip between days 3 and 7 corresponds to a period when observers were beginning to identify subtle call notes. Also, the training site was changed, distance estimates were added to the count on day 7, and training was interrupted by a weekend, all factors that tend to reduce comparability.

In a random selection of 50 paired station counts from the 1298 surveyed during the 1980 Maui Forest Survey, only 8 (16%) had identical lists (Fig. 1). This low figure resulted from a total number of species recorded per station that ranged from only three to nine ($\bar{x}$ = 5.6) in the simple Hawaiian ecosystems. The sample was drawn from a variety of dry and wet forest habitats. We wished to look at variability within a single prime forest ecosystem, and chose upper elevation (1372 to 1982 m) ohia (*Metrosideros collina*) forest inhabited by the Crested Honeycreeper—the same system in which our final training session was held. Thus we can directly compare survey results with training day 12. Four transects were randomly selected from the 13 that crossed this forest. Forty-five stations sampled the habitat; on 11 (24%) of them the paired observers had identical species lists (Fig. 1). The large difference between this and train-

TABLE 1
POOLED DATA FOR EIGHT 5-MINUTE COUNT PERIODS FROM DAY 12, MAUI SURVEY TRAINING SESSION, 20 MAY 1980. GROUPS A AND B WORKED INDEPENDENTLY IN SEPARATE AREAS, AND CAN NOT BE DIRECTLY COMPARED

	Group A Observer			Group B Observer		
Species	4	3	5	2	6	7
Amakihi[a]	10	8	12	17	16	18
Maui Creeper	25	24	31	31	27	28
Maui Parrotbill	2	0	1	0	1	0
Apapane	64	65	61	80	88	81
Crested Honeycreeper	23	26	17	22	21	24
Iiwi	16	11	9	22	18	15
Red-billed Leiothrix	5	4	3	9	12	8
Japanese White-eye	13	8	5	7	9	6

[a] See Table 2 for scientific names.

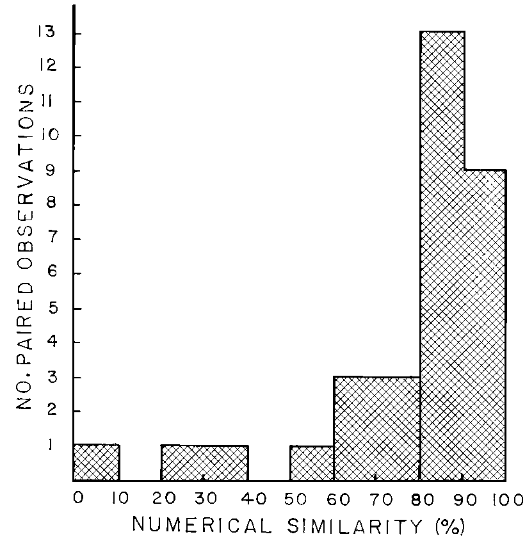

FIGURE 3. Numerical similarity between paired observers in the number of individuals recorded per species during the Maui Forest Bird Survey, 1980. The data set is presented in Table 2.

ing day 12 (46% identical lists) is largely a measure of the effect of the 18.4 m distance separating observers under actual survey conditions.

CONGRUITY OF LISTED SPECIES

A much clearer view of observer variability is provided by comparing the frequency at which species are recorded in common. During the training session this ranged from 67% (day 1) to 86% (day 12); daily sample sizes (number of species recorded, station pooled) ranged from 131 to 541, and totalled 1767. Of these, 1330 (75%) were simultaneously recorded by both observers. The improvement shown from day 1 to day 12 is significant ($\chi^2 = 34.56$, $P < 0.001$).

During the 50 randomly selected survey counts, species were recorded 280 times: observers found 189 (67%) of them during the same count. The remaining 91 were recorded by only one observer during a count period. Within the ohia forest, 10 species were recorded 289 times, 226 (78%) by both observers during each station count.

DECREASING VARIABILITY BY INCREASING COVERAGE

The above examples indicate that even in simple systems it is unrealistic to expect observers to obtain identical species lists. This is, however, far less important than the overall congruity between observers for listed species (67% in all habitats, 78% in ohia forest), and the relationship between species lists over repeated samples. By increasing sampling coverage the apparent differences, as reflected in a compari-

TABLE 2
SURVEY RESULTS FOR PAIRED OBSERVERS IN OHIA FOREST ON FOUR RANDOMLY SELECTED TRANSECTS, MAUI FOREST BIRD SURVEY, 1980. THE NUMBER OF STATIONS SAMPLED ON EACH TRANSECT WERE: TR. 3 (12), TR. 9 (9), TR. 10 (13), AND TR. 18 (11)

	Tr. 3 Observer		Tr. 9 Observer		Tr. 10 Observer		Tr. 18 Observer	
Species	1	2	3	4	3	1	4	5
Amakihi, *Loxops virens*	22	23	25	22	18	17	15	9
Maui Creeper, *Loxops maculata*	48	52	20	25	21	26	4	0
Maui Parrotbill, *Pseudonestor xanthophyrs*	0	0	1	1	3	2	0	0
Apapane, *Himatione sanguinea*	38	34	26	22	31	37	27	33
Crested Honeycreeper, *Palmeria dolei*	52	51	10	10	10	14	32	33
Iiwi, *Vestiaria coccinea*	10	10	25	31	5	22	18	13
Melodius Laughing-thrush, *Garrulax canorus*	0	0	0	0	0	0	1	3
Red-billed Leiothrix, *Leiothrix lutea*	65	40	23	20	28	33	17	14
Japanese White-eye, *Zosterops japonica*	23	19	11	13	5	9	15	12
Cardinal, *Cardinalis cardinalis*	0	0	6	6	0	0	0	0

son of single counts, decrease dramatically. In the 10-species ohia forest system, observers on each of the four transects recorded identical species when all counts within the habitat (from 9 to 13 per transect) were summed, with a single exception. In this instance one observer recorded a Maui Creeper (*Loxops maculatus*) missed by the other. Rare birds do, obviously, increase the likelihood of difference.

In Figure 2 we present data illustrating the number of samples needed to achieve identical species lists (excluding the creeper). The number of station counts are of adjacent pairs, triplets, etc. of stations along each transect. For example, three station comparisons include stations 1–3, 2–4, 3–5, etc. for each transect. Clearly, increasing coverage rapidly improves congruity between observers. It is a relatively simple matter to test for this effect under any set of habitat or species richness conditions, and such a test should be an integral part of one's experimental design. In the Maui example, a minimum of five counts per transect in prime habitat effectively overcomes much of the inherent variation between observers, even when they stand relatively far apart (18.3 m). We must stress, however, that merely increasing sample size is no substitute for a training program, and that coverage or sampling intensity may have to be increased considerably to detect rare or elusive species.

Numerical Differences among Observers

Bird observers detect differing numbers of individuals during their surveys, a fact that has provoked considerable concern (Berthold 1976). Fortunately, these numerical differences can be reduced by training. During the 5-minute simultaneous counts on training day 1, fully 41% of all species comparisons between observer pairs revealed differences of greater than 50%. By day 12 variation of this magnitude had decreased to 13%. During the same period the frequency of paired observations showing less than 20% difference increased from 26% to 52%. The numerical similarity between observers was impressive (Table 1), particularly for the common species. The highest percentage differences between observers generally occur with the rarer species.

We have 32 sets of paired observations for the Maui survey in ohia forest (Table 2), with from 2 to 105 individuals of each species recorded by the two observers. The majority (69%) of these sets agree within 80% (Fig. 3). At the low end, observers recorded a difference of greater than 60% for only three species. Important here is that two of these species, rare on their transects, were only found four times, and one of them (Maui Creeper) was not detected by one observer. There are times, however, when observer differences are extreme for relatively commonly encountered species (see Transect 10, Iiwi). Fortunately, such differences have been relatively rare (3% in this sample) in Hawaii, and their rarity is in part a function of the intensive training program and, perhaps, the simple fauna. Such differences also have another important cause, and this generally traces to an inequality in hearing acuity between observers. Thus an observer with "good ears" is sampling a larger area by hearing more distant birds. For the Iiwi on Transect 10, the median detection distance for Observer 1 was 58 m, while that for Observer 3 was only 22.9 m. Thus although Observer 3 only saw 23% of the birds recorded by Observer 1, he was sampling only 16% of the area. The effect of these differences in numbers is largely moderated when bird densities are calculated from distance estimates.

DISCUSSION

That differences exist among observers is indisputable, and they must be seriously addressed in any bird sampling program. Although many physical and psychological factors can not be eliminated, their impact can be greatly reduced by an extensive program that begins before observers are selected. Careful initial screening of applicants can eliminate the more obvious visual, aural, and psychological factors that increase observer variability. A rigorous observer training program further reduces inherent variation, but does not eliminate it. Such training must place heavy emphasis on distance estimation, for the ability to estimate distances accurately offsets inherent differences in aural and visual acuity when variable circle counts or certain linear transect counts are used. We must remain aware that observers are not perfect nor identical, and that we must seriously address and decrease this source of variation in order to minimize its impact upon our data, and hence our understanding of those population parameters we are attempting to evaluate. The reliability of "experienced" observers can be improved by pairing them and conducting simultaneous counts and thereby identifying problem species. Elimination of problem species for experienced observers might only take a day or two of work with audio and visual aids.

ACKNOWLEDGMENTS

The development of the training methods for the Hawaii Forest Bird Survey evolved over a five-year period. Its present format is the result of input from a large number of people. We particularly thank Philip Ashman, Tim Burr, Tonnie L. C. Casey, Jack Jeffery, and Peter Pyle for their valuable comments. Fred L. Ramsey has provided insightful comments and inspiration during the entire period.