BEHAVIORAL SAMPLING METHODS FOR CETACEANS: A REVIEW AND CRITIQUE

JANET MANN

Department of Psychology and Department of Biology
Georgetown University
Washington, DC 20057, U.S.A.
E-mail: mannj2@gunet.georgetown.edu

ABSTRACT

Behavioral scientists have developed methods for sampling behavior in order to reduce observational biases and to facilitate comparisons between studies. A review of 74 cetacean behavioral field studies published from 1989 to 1995 in *Marine Mammal Science* and *The Canadian Journal of Zoology* suggests that cetacean researchers have not made optimal use of available methodology. The survey revealed that a large proportion of studies did not use reliable sampling methods. *Ad libitum* sampling was used most often (59%). When anecdotal studies were excluded, 45% of 53 behavioral studies used *ad libitum* as the predominant method. Other sampling methods were continuous, one-zero, incident, point, sequence, or scan sampling. Recommendations for sampling methods are made, depending on identifiability of animals, group sizes, dive durations, and change in group membership.

Key words: methods, sampling, observations, behavior, vocalization, Cetacea, ethology, quantitative.

Researchers studying the behavior of cetaceans at sea face several unusual methodological challenges. Many cetaceans swim rapidly, range over long distances on a daily basis, and have seasonal migrations of thousands of kilometers. Cetaceans are difficult to follow because they disappear during dives and do not leave long-lasting traces, such as tracks, scats, or dens.

To meet these challenges, cetacean researchers have developed photoidentification methods and some clever technical solutions. For example, individuals of many species can be identified by natural markings, allowing researchers to link sightings separated by years and thousands of kilometers (Hammond *et al.* 1990). Individual animals can be followed in more detail by tagging them with radio, acoustic, or satellite tags. Diving behavior can be monitored using time-depth recorders. Hydrophones and underwater video cameras can provide continuous data on the behavior and communication signals of animals for shorter periods of time.

In spite of these impressive technical advances, methods for sampling ce-

tacean behavior have received less attention. In this paper, I review quantitative observational methods typically used in cetacean studies and offer specific sampling suggestions.

Altmann (1974) points out that most field observations of behavior involve sampling decisions, whether those decisions are explicit or not.

> "*Sampling decisions are made whenever the student of social behavior cannot continuously observe and record all of the behavior of all of the members of a social group. . . . We suspect that the investigator often chooses a sampling procedure without being aware that he is making a choice. Of course, he does not thereby escape from the consequences of that choice.*" (Altmann 1974:229).

The lack of an explicit protocol for sampling was termed *ad libitum* sampling by Altmann (1974), and she pointed out that it typically entails scoring "as much as one can" or whatever is most readily observable of the behavior of a group (Altmann 1974:235–236). These kinds of observations may be necessary as one learns to identify behaviors, as one plans a study, or as one observes rare but significant events. However, *ad libitum* observations suffer from a variety of potential biases. Different individuals may be more or less visible. Some behaviors may be more salient and more readily recorded than others. Individual animals may alter their behavior depending on how visible they are to other animals (and to observers). The same observer may concentrate on different behaviors during different observation periods, and different observers may notice or attend to entirely different behaviors during the same observation period. Such biases indicate that *ad libitum* data are not appropriate for estimating rates of behaviors or for comparing rates across subjects or across studies. The selection and appropriate use of sampling methods that yield unbiased estimates of behavior are critical to the scientific validity of any study.

In this paper, first I review the methods employed in the majority of recent cetacean field studies, including sampling method and protocols, and discuss the general advantages and pitfalls of each sampling method. Second, I recommend specific sampling methods for cetaceans, depending on identifiability of animals, group sizes, dive durations, and change in group membership.

Literature Survey

To evaluate what methods are currently used by cetologists, I surveyed papers on cetacean behavior published from 1989 through 1995 (see Appendix 1 for details). To limit the survey, I selected two peer-reviewed journals that together published the majority of studies in wild cetacean behavior, *The Canadian Journal of Zoology* (CJZ, 31% of studies) and *Marine Mammal Science* (MMS, 29% of studies).

Seventy-four studies were reviewed, 38 in CJZ and 36 in MMS. The following information was noted: species, age and sex classes, number of animals, number of observation or recording hours, group-size minima and maxima, whether animals were individually identified, types of behaviors recorded (including vocalizations), definition of group and definition of behaviors, and

methods of observation and sampling. In many cases, essential information in these categories was not provided (Appendix 1).

CRITIQUE OF METHODS

For every behavioral study, two basic kinds of sampling decisions must be made. One choice concerns which subject(s) one watches and for how long; the other concerns the details of how behavior is recorded (Martin and Bateson 1986). I will discriminate these as "follow protocol" and the "sampling method." "Follow protocol" refers to how long an observation extends and to whether researchers follow a group or an individual animal. "Sampling method" refers to the procedures used to sample the behavior of individuals or groups. Such a distinction is necessary because the sources of error or bias differ according to each protocol and according to each method. For example, in the ethological literature the term "focal-animal sampling" is used to refer to data collection that involves sampling of behavior of one individual for a set period of time. However, there are really two separate components, the follow protocol (following an individual) and a sampling method (which could be continuous sampling, point sampling, *etc.*). To state only that one's method is "focal-animal sampling" would be insufficient. For example, a researcher could collect data on a focal individual systematically at regular intervals or opportunistically (*ad libitum*) by irregularly noting behaviors of interest.

In the following sections, the prevalence and the costs and benefits of different follow protocols and sampling methods are discussed. Table 1 is a guide, illustrating how different techniques (follow protocol and sampling method) are likely or unlikely to be effective, depending on the characteristics of the animals under study. My use of the term "identifiable" in Table 1 refers to cases when observers can identify known individuals (such as in a longitudinal study) or can discriminate between animals sufficiently to keep track of the same animal. When I refer to "group," both for group size and group membership, this concerns the number of animals close enough together to be potentially confused with each other.

Follow Protocol

For the studies surveyed, five different follow protocols can be defined: *survey, group-follow, individual-follow, tracking,* and *anecdote.* Three studies used more than one protocol (*e.g.,* group-follows and surveys). Two studies did not provide enough information to allow classification by protocol.

Survey refers to encountering groups or individual animals and staying with those animals for brief periods to census, for example, the number of animals, identifications, location, and behavior. If observers typically monitor groups for 30 min or less, then their protocol is identified here as "survey." Surveys comprised 16% ($n = 12$) of the studies reviewed. Surveys provide a "snapshot" of animal life and are very useful for tracking patterns of association and for analyzing demographic, reproductive, and ecological factors. Surveys are par-

Table 1. Recommended uses for different sampling methods depending on subject characteristics.

Method	Rapidly identifiable?		Group size		Group membership rate of change		Dive time	
	yes	no*	<10	≥10*	<12 h*	≥12 h	<10 min	≥10 min
Individual-follow protocol								
Continuous	yes	no	yes	?	yes	yes	yes	?
Point∧	yes	no	yes	?	yes	yes	yes	?
PAS	yes	no	yes	?	yes	yes	yes	?
Group-follow protocol								
Scan	yes	yes	yes	yes	?	yes	yes	?
Sequence+~	yes	?	yes	?	yes	yes	yes	yes
Incident+	yes	yes	yes	?	?	yes	yes	yes

This table is a guide illustrating how well particular sampling methods are expected to fare depending on behavior and characteristics of subjects under study. Numeric categories for group size, rate of group membership change, and dive time are somewhat arbitrary. Familiarity with each species is necessary to determine what sampling method might be most effective. ∧— recommended for spatial proximity measures and behaviors that are rapidly identifiable; +—recommended for infrequent and obvious behaviors, such as breaches or lobtails; ~—sequence parameters must be carefully defined; *—more challenging sampling conditions; yes = can probably use sampling method if no other problems; no = unlikely this method will work unless animals are solitary; ? = may be possible to use sampling method, depending on other conditions.

ticularly valuable for addressing population-level questions (*e.g.*, density and distribution of animals). Nearly every study of wild cetacean behavior relies on surveys for basic population information.

If observers monitor groups of animals for longer than 30 min, their protocol is classified as *group-follow.* The group-follow was the most common protocol used (43%, $n = 32$) in the studies reviewed. Although following a group is easier than following an individual, rigorous systematic sampling can prove to be more difficult, primarily because the observer's attention is naturally drawn to more obvious behaviors or more visible animals. The advantages of group-follows are: (1) many individuals can be sampled, and (2) some questions on the temporal and spatial scale of social structure can be examined (see Whitehead 1995, 1997). If cetologists follow groups, the behaviors or individuals sampled must be limited to those which can be reliably or consistently recorded. Researchers using the group-follow protocol must consider how group composition changes may bias data collection. The sampling protocol must include a "decision rule" for when one or more animals leave the group. One possible solution is to alternate between "stay with the smaller group" and "stay with the larger group." Without such "rules," observers may be biasing their sample with the behavior of large groups, active groups, or other attributes that affect the observer's decisions on which group to sample.

Appropriate sampling techniques applicable during group-follows are discussed below (see Table 1).

If observers monitor an individual regardless of whether the animal is in a group or not, their protocol is classified as *individual-follow*. Only 12% of the studies reviewed used this protocol. The *individual-follow* is roughly equivalent to focal-animal sampling. The critical feature of this method is to focus on one animal and systematically record behavior that is defined *a priori*. When sampling individual behavior, researchers may still collect *ad libitum* data on other animals or scan the group at regular intervals to determine group membership. However, with few exceptions, data collection on non-focal animals should not compromise focal data.

The individual-follow enables the observer to focus on the individual animal's "perspective." What is the day in the life of that animal like? Who do they approach, avoid, stay close to, interact with, mate with, and fight with? Observing the continuous stream of individual behavior in different contexts is central to the understanding of the dynamics of social relationships. Time-budget data for individual animals are more appropriate for most studies than group-activity data because individuals of varying age and sex have different behavioral and ecological strategies. Both methodological and theoretical arguments support focusing on the individual as the unit of analysis. Selection operates at the level of the individual (Williams 1966), and this perspective is critical to understanding the behavioral and reproductive strategies of cetaceans. Group-living cetaceans may rely on each other for survival and reproduction, but the costs and benefits of group living are unlikely to be shared equally among all group members. Measures of such variation are particularly relevant to the study of how natural and sexual selection has shaped the evolution of social systems and behavior patterns in a species.

The success of systematic sampling with the individual-follow depends on how rapidly animals can be identified or discriminated from other group members when they surface (Table 1). Several factors influence this, including the size and stability of groups and the duration of dives or "out-of-sight" periods. The optimal situation for individual-follows is that of a small stable group where individuals can be identified as soon as they come into view. If individuals are not readily identifiable upon surfacing, individual-follows may still be used if the animals are solitary much of the time (*e.g.*, humpback whales). Group size is also a factor. No matter how identifiable an animal is, if there are hundreds of animals, it will be hard to keep track of a focal animal (Table 1). Brief focal samples, spanning 5 min or less (one or two surfacing bouts), may be appropriate under such conditions. Individual-follows might be facilitated by focusing on a readily identifiable member (*i.e.*, based on size, marks, tags, or radiotracking), although the potential biases inherent to such selection must be considered.

Although sampling decisions depend on the observation conditions, long individual focal-follows are highly recommended where possible. Rogosa and Ghandour (1991) have argued (mathematically) that short samples are one of the biggest sources of unreliability in sampling uncommon behaviors. Long

individual follows may also be practical, given the search-time costs of sampling different individuals on the same day. If the researcher is likely to lose track of the focal animal in certain behavioral contexts (*e.g.*, foraging), then scans or other sampling techniques may be applied to prorate or correct for such biases. Individual follows often enable the observer to identify the sources of sampling bias (*e.g.*, conditions that prove difficult for monitoring individual behavior). If the study design requires following individuals for long periods of time, but individuals are not rapidly identifiable, then visual, radio, or acoustic tags may be necessary.

Tracking refers to studies that electronically monitor individuals' locations or behavior (through hydrophones, transponder tags, or other devices). Tracking is particularly valuable if researchers need to continuously record the behavior of an animal over long periods. Six studies (8%) reviewed here tracked small numbers of animals using radio and transponder tags to record diving and vocal behavior. This approach can be relatively expensive, and the attachment process or device may affect behavior. Sample sizes tend to be small in tracking studies.

An *anecdote* is a descriptive report of a single event or series of events, such as birth or predation. Twenty-eight percent ($n = 21$) of the studies analyzed here were classified as anecdotes. Anecdotes are a valuable means of describing rarely observed events.

The survey and group-follow may be considered "group-protocols"—where groups are monitored unless animals are alone. The individual-follow is an "individual-protocol"—where a specific individual is monitored regardless of whether or not it is a member of a group. Tracking can involve groups or individuals. The merits of different protocols must be weighed against their impact or influence on the animal's behavior. Few data are available on whether or how tagging, equipment noise (*e.g.*, engine, tag, depth sounder), and prolonged *vs.* brief follows directly affect the behavior of cetaceans. Indirect effects may also occur by alteration of the behavior of cetacean prey or predators. By combining different approaches, or directly testing for the effects of these approaches, observers can find ways to minimize their impact on behavior.

Sampling Methods

Given that a researcher has chosen a protocol for following the animals, a number of sampling methods may be used. These include *ad libitum, continuous, focal-group, one-zero, point, scan, predominant activity, sequence*, and *all-event/ incident sampling* (defined below). The success of each sampling method depends, in part, on whether a researcher is focusing on *events* (brief behaviors, measured in frequency) or *states* (long behaviors of measurable duration). Although the distinction is convenient, events and states are on a continuum (*i.e.*, all events have durations and all states have frequencies).

Ad libitum sampling—*Ad libitum* samples are "typical field notes" (Altmann 1974). The observer writes down what seems of interest. *Ad libitum* sampling does not involve systematic constraints on what is recorded and when it is

recorded. The initial phases of behavioral study often involve some *ad libitum* sampling in order to delineate and define behaviors and research questions. Most observers continue to collect some *ad libitum* data throughout the course of their study. In the survey of studies, the sampling method was classified as *ad libitum* only if it was the predominant method used. *Ad libitum* sampling was the most common technique used (59% of studies). *Ad libitum* sampling is useful for certain kinds of comparisons, but not for estimating rates of behavior or for comparing behavior patterns of different age or sex classes (Altmann 1974). As Altmann points out, some animals may be more notice-able, either because of their behavior (or activity level), size, or level of habit-uation. *Ad libitum* sampling may be applied during individual focal-follows (*e.g.,* prey captures by focal animal) or group-follows (*e.g.,* displays). *Ad libitum* data are valuable for looking at the direction of interactions or at what happens within a dyad if the direction of the interaction is unlikely to be influenced by the probability of seeing it. For example, *ad libitum* data are commonly used in ethology to assess dominance relationships, where one animal "wins," and the other "loses" an agonistic interaction (*e.g.,* see Samuels and Gifford 1997). Rates of agonistic interactions cannot be acquired through *ad libitum* sampling, because the likelihood of seeing an interaction may be affected by the rank, size, sex, or other characteristics of the animals (*e.g.,* conflicts between adult males are more likely to escalate and thus be observed than conflicts between adult females). However, given that two animals are fighting, the outcome of the fight is unlikely to influence or be influenced by the probability of seeing the interaction. *Ad libitum* sampling is effective for examining di-rected displays, in which both signaller(s) and recipient(s) can be determined. Sociomatrices are an excellent means of illustrating some patterns of interac-tion with *ad libitum* data. A sociomatrix has all potential interactants listed on the top and along the side, with one axis indicating actor or "winner" and the other as recipient or "loser." Frequencies are tallied within each cell of interactants. When dominance hierarchies are linear, for example, then most entries will fall on one side of the diagonal. Sociomatrices can be applied to other behaviors sampled *ad libitum.* Mann and Smuts (in press) recorded 1,311 rubbing events between nine wild *Tursiops* newborns and their mothers during focal individual-follows (189 h). Rubbing rate could not be determined, nor whether some infants rubbed more than others. But rubbings were extremely asymmetrical, with infants initiating and performing 99% of the rubbing on their mothers.

 Ad libitum data are a valuable part of any field study but should not be represented as rates, proportions, frequencies, or other unbiased estimates of behavior. The use of *ad libitum* data for anecdotes is perfectly appropriate. Rare events such as predation, a birth, or a lethal fight, can provide critical infor-mation and insights, and *ad libitum* sampling is likely to be the only way such events are recorded.

 Continuous sampling—Alternate names: event sampling, frequency sampling, focal-animal sampling. Continuous sampling is a systematic record of fre-quencies or durations for a specified set of behaviors. Researchers occasionally

use the term "focal-animal sampling," to indicate continuous sampling of behavior during an individual-follow. Measuring the exact times (and durations for behavioral states) of every occurrence of a behavior is very demanding for the observer. Scoring event frequencies within time blocks can simplify data collection. The reliability of continuous data can be easily compromised if the observer tries to record too many behaviors at once, especially when the animals are active. Thus, Altmann (1974) recommends that this method be used for one or two animals at most.

Continuous data on associations, diving behavior (with transponder tags), other behaviors, or vocalizations were collected in 14% ($n = 10$) of the studies surveyed. In four of these studies researchers observed animals directly. Two studies used tags to track individuals. The remaining four collected continuous data on small groups of animals using videotape or multiple observers.

Continuous sampling of behavior is relatively simple for activities at the surface (during surfacing bouts). Surfacing-bout durations, breath frequency, dive types, surface-display rates, and synchronous surfacings of the focal animal and others may be recorded on a continuous basis for many dolphin and whale species. Even deep-diving animals, such as sperm whales, may have prolonged bouts of socializing at the surface (Whitehead and Weilgart 1991). If the activities are brief, or difficult to time, then they can be scored as events (frequencies), rather than states (onset to offset). Records of behavior sequences are typically intact in continuous sampling. As Table 1 indicates, continuous sampling is most likely to succeed when animals are rapidly identifiable, live in small groups, and dive for short periods. Under these sampling conditions, observers can keep track of a specific individual for long periods of time. Continuous data are the richest source of information on social behavior and relationships, because such data include information on details, sequences, actors and recipients, rates and durations of behavior for individual animals.

While it is simple to record vocalizations continuously using a tape recorder, for these data to qualify as continuous data for an individual follow, the observer must be able to identify which vocalization comes from the focal individual. Use of tags that can record or transmit acoustic data is one technique for achieving this (*e.g.,* Tyack 1985, Tyack and Recchia 1991). Another technique involves passive acoustic localization (*e.g.* Clark 1980, Freitag and Tyack 1993, Frankel *et al.* 1995). Fusion of these acoustic and visual data remains a challenge to anyone interested in cetacean communication. As with behavioral studies, the study of cetacean vocalizations still hinges upon such basic information as the species, sex, or age class of the animal producing the call. The study of how these sounds are used in social communication will require increased application of methods to identify which animal produces which sound during interactions within and between groups.

Focal group sampling—Alternate names: focal subgroup sampling, group sampling, predominant group activity sampling. Focal-group sampling is a continuous assessment of group activity. Group activity may be scored at intervals (*e.g.,* indicate predominant group activity every 5 min) or continuously (*e.g.,* the group rested from 1122–1147). This method is widely used in ce-

tacean research when observers follow groups, and Altmann (1974) has been cited in support of it (e.g., Shane 1990). Altmann (1974) used the term "focal sub-group sampling" to refer to continuous sampling with more than one individual. However, Altmann described focal sub-group sampling as appropriate only under a very restricted set of conditions "in which all individuals in the sample group are continuously visible throughout the sample period . . . if one is working with observational conditions that are less than perfect, focal-animal sampling should be done on just one focal individual at a time, or at most, a pair (e.g., mother and young infant)" (Altmann 1974:243–244). Twenty-seven percent of the studies surveyed here (45% of ad libitum studies) used "focal group sampling." These studies scored "group behavior." Dozens of animals, not pairs, were sampled using focal-group sampling. I chose to classify these studies as ad libitum because none of the conditions necessary for unbiased focal-group-sampling were met. That is, observers could not have continuously observed all animals equally in a group regardless of activity. Some observers appeared to visually assess group activity, perhaps by informally scanning the group. However, this method is not explicit enough in how subjects were sampled and does not differ in any substantial way from ad libitum sampling.

A few studies identified their method as predominant group-activity sampling. This should not be confused with predominant activity sampling (defined below). Predominant group-activity sampling is essentially the same as "focal group sampling" and was coded as ad libitum in the survey. During predominant group-activity sampling, the observer defines group activity based on an assessment of what most (>50%) of the group is engaged in over an interval, focusing on the *proportion of individuals* estimated to be engaging in a behavior, *not the proportion of time an individual* engages in a behavior for that interval. An estimate of predominant group activity *can* be achieved by explicitly scan sampling over 50% of the individuals in the group, rather than by "watching the group."

Altmann (1974) prescribed a restricted set of conditions for focal sub-group sampling because of the many biases inherent to watching groups. First and foremost, the observer's attention is naturally drawn to animals that are most visible, most active, or most interesting. No observer can possibly continuously track and record all behavior of all individuals simultaneously under all conditions; it is difficult enough to record the behavior of one animal. Group activity may be determined if all members of a small group are cohesive and engaging in the same behavior, such as resting closely at the surface, but it may be difficult to determine if some animals are engaging in other activities. The accuracy of focal-group sampling is dependent upon group size, cohesiveness, and activities of the animals, thus potentially introducing biases into data collection. Some behaviors, such as socializing, may be particularly obvious to observers, and even though most of the animals are resting, one might score the more visible behaviors as group activity. While using group-sampling, observers are often making implicit assumptions about the relative im-

portance of behaviors of different age and sex classes. If two mothers are hunting, but their calves are traveling, should one call it hunting or traveling?

With video, continuous sampling of groups is possible because observers can rewind the tape and code the behavior of different individuals separately (thus making the process equivalent to conducting multiple continuous focal-animal samples simultaneously). However, simultaneous behaviors of different individuals within the same group are likely to depend upon one another, which must be taken into account for statistical analyses. Focal-group sampling is not only used during the group-follow protocol, it is also used during surveys to identify "group activity." However, this too is subject to the same criticisms.

One study (Fragaszy et al. 1992) has compared focal-group sampling with focal-individual point sampling and scan sampling (see definitions below). This study, although on primates, is particularly relevant for cetacean studies because, like cetaceans, capuchins and squirrel monkeys move in and out of view and are difficult to follow for continuous periods of time. Fragaszy et al. (1992) compared foraging time budgets of squirrel monkeys using both focal-group sampling and focal-animal point sampling methods and concluded that focal-group sampling overestimated foraging considerably. They came to this conclusion by correlating individual-focal point sampling rates with focal-group sampling rates. The focal-group sampling error rate (computed by JM) ranged between 39% and 63%. They also compared scan sampling of capuchin monkey groups to focal-animal point sampling. Scan sampling fared much better. When time budgets for scan sampling and focal-animal point sampling for foraging among capuchins were compared, the error rate was only 0.8% (computed by JM). For other, less frequent behavioral states, the error rates for scan sampling ranged from 10% to 26%. Focal-group sampling error rates for other behavioral states were not compared. However, error rates are lowest for more prevalent behaviors, and foraging was the most common state. Thus, error rates are likely to increase for other states. Such error rates should give pause to anyone considering group-sampling.

One-zero sampling—Alternate names: time-sampling, Hansen frequencies, interval sampling, partial interval sampling, method of repeated short samples, and modified frequencies. One-zero sampling entails scoring whether or not a behavior occurs during an interval (*e.g.,* 30 sec), rather than scoring how frequently or how long the behavior occurred. For example, an observer may score whether or not an animal vocalized during 30 seconds, or whether or not it rubbed, rather than the frequency of vocalization events or the duration of rubbing (a state). Nine percent of the studies surveyed employed one-zero sampling. With few exceptions, this technique is not recommended, because one-zero scores do not represent frequency or duration and are prone to very high rates of sampling error (see Mann et al. 1991). Although one-zero sampling has no recommended uses (Altmann 1974), there are some natural types of one-zero data (categorical variables) that are not represented by the flow of continuous behavior but may occur over very long intervals. For example, if group fissions and fusions are rare, then whether or not an animal joins a

group during a day or field season is a significant piece of information and may be more biologically meaningful than the absolute rate. However, one must already know something about the rates and durations of behaviors of interest before employing one-zero sampling.

Researchers sometimes use one-zero events to define states. That is, one-zero data may assist in the development of an ethogram but not be part of the sampling method. For example, if two animals are seen in contact during an interval, this may "define" socializing regardless of the duration or frequency of contact. Or, one might define foraging based on whether or not the animal flukes-up at the dive. It is important then to distinguish between the use of one-zero (categorical) information to help define behaviors that are presumed to be continuous and treating one-zero scores as state behaviors themselves.

Point sampling—Alternate names: instantaneous sampling, on-the-dot sampling, time sampling. *Point sampling* entails scoring activity as a "snapshot" at a given moment (*e.g.*, every 30 sec). The observer may score states, such as distance to others, activity, or other information on a point-sampling basis. Five percent of the studies surveyed used point sampling. It is a reliable method that is widely applied in ethological studies but has been rarely used for cetaceans. One possible explanation is that the "point" often happens when the animal is submerged. The few studies reviewed here that used point sampling did so by electronically tracking the animal. For direct observations, point sampling can still be applied with two basic strategies. (1) The animal's behavior is sampled at the first surfacing after the interval point. Two factors may lead to sampling biases. One ends up sampling surface behavior rather than the behavior occurring underwater at the point. By examining the relationships between surface and subsurface activities, these biases may be minimized. Another bias is that the behavior is not sampled at regular intervals, but at intervals determined in part by the subject's surfacing and diving behavior. If the observer sampled at 1-min intervals, and animals tended to remain close to the surface during socializing and dive deeply for longer periods during foraging, then the observer could successfully sample socializing but would miss several sampling intervals during foraging. Social behavior would be overrepresented in relation to foraging. This problem can be alleviated by establishing a point interval greater than the typical long dives of the subject. (2) The second strategy is to use point sampling for behavioral states *if one can assume that the state continues when the focal animal moves in and out of view.* Observers often infer behavioral state based on surfacing pattern, proximity of other individuals, and other cues (*e.g.*, flukeprints or bubbles at the surface). The observer may apply a convention such as scoring a behavioral state at the "point" if the animal engaged in the behavior prior to diving or disappearing from view and was engaged in the same behavior upon surfacing. This convention may be applied if the "out-of-sight" periods tend to be much shorter than the bout duration of the behavioral state. Another convention is to score the behavior as that last observed, or, alternatively, first seen upon surfacing. Such conventions must be explicitly defined, justified, and the po-

tential biases considered. As Table 1 suggests, point sampling can be applied under many conditions but becomes problematic if dive times or out-of-sight periods are long and group sizes are large (making it difficult to keep track of an individual animal).

Point sampling can be a very useful method to determine time budgets or diurnal patterns of behavior. This method is also a useful supplement to other focal-individual sampling methods. For example, the observer may record an animal's speed at 5-min intervals, or note a focal animal's nearest neighbor based on who surfaces closest to the focal animal at the first surfacing after each 10-min interval. Although point sampling is typically recommended for behaviors of appreciable duration, activity can be more difficult to score on a point-sampling basis. In reality, even with the animal in full view, it may take several seconds or more to "decide" what an animal is doing. It is important that the observer decide as quickly as possible, so she/he does not inadvertently wait until the animal does something that is easier or more interesting to score. Brief behaviors or events are difficult to observe accurately with point sampling because such behaviors are often missed at the "point."

Scan sampling—Scan sampling entails taking a "point" or "instantaneous" sample of an individual's behavior or location before moving to the next animal and doing the same. Scans are conducted either at regular intervals (*e.g.,* sample each animal at 10-sec intervals), or as quickly as possible (*i.e.,* search for the next animal as soon as the last was sampled). Both point and scan samples are good techniques for measuring states, but brief events are likely to be missed. Three percent of the cetacean field studies reviewed here used scan sampling.

Scan sampling is very valuable for sampling behavior when focal-individual observations are not possible or desirable, or if the researcher wishes to keep track of group activities. An explicit scan-sampling technique should also be considered for surveys of groups, rather than using *ad libitum* sampling to identify "group activity." Observers sometimes use a random scan-sampling schedule, or just scan from one side of a group to the other. Alternatively, if group sizes tend to be large and the animals move rapidly, but age or sex classes are distinctive in size or markings, then the researcher may scan age and sex classes separately (*e.g.,* scan large males, then females with dependent calves, then immature animals). This would leave some age/sex classes unsampled but would enable the observer to scan a group of fast-moving animals more effectively.

Scan sampling and point sampling share practical difficulties: the time it takes to decide the animal's activity and out-of-sight or diving periods. One may employ the same strategies for scan sampling as for point sampling, but with scan sampling the observer samples each individual only once per session. In some cases, the observer may need to watch each individual for 3 min or more (approximating a short individual-follow) but uses the midpoint (*e.g.,* minute 2) to define the behavioral state. Thus, one data point is taken for each animal in succession, although each observation period is longer than for a typical scan. The same conventions used for point-sampling out-of-sight periods can be adapted to scan sampling.

Potential uses for scan sampling are varied (see Table 1). First, an observer may scan a group at fixed intervals (*e.g.,* every 10 min) to assess which is still present, which is close to a focal animal, *etc.* Second, an observer may scan a group to rapidly estimate group activity by identifying each animal's activity as it surfaces. This is only possible for groups where one can keep track of each individual within the group during surfacing and depends in part on the degree of interindividual overlap in surfacing bouts. The observer may be able to scan rapidly enough to minimize the likelihood of resampling the same individual or scan from the front to the back of a group if direction changes are infrequent. If the scan procedure inadvertently allows for repeat sampling within a session, then the resampling of highly identifiable individuals would alert the observer to this problem. Third, an observer might use it to assess nearest neighbors for each animal. Fourth, one could record both activity and nearest neighbor for each group member. If group scans are conducted during individual-follows, then it is important that the scans can be completed very quickly to prevent the observer from being distracted from the focal individual. With scan sampling, cetacean researchers can broaden their dataset to look at coordination of group activities and refined measures of association for a number of animals of different age and sex classes.

Predominant activity sampling—Predominant Activity Sampling (PAS), developed by Hutt and Hutt (1970), refers to scoring individual behavior as the predominant activity over some interval (*e.g.,* 30 sec), only if that behavior occupied 50% or more of that interval (>15 sec). This method is only useful for measuring states. No studies in the survey used PAS. It is an empirically valid technique for estimating the proportion of time during which behavioral states occur (Tyler 1979). Very brief behaviors or displays (events) will not be represented in PAS data unless the observer uses very short intervals. It is important that the interval length is brief enough to capture the briefest states of interest.

PAS is useful for sampling animals that go in and out of view for brief periods. In a sample of *Tursiops* calves, some types of data have been collected using both focal individual point samples and focal PAS simultaneously (J. Mann, unpublished data). Thus calf activity budgets could be compared based on sampling type. For example, "calf position swimming" (when the calf is in contact under the mother) was measured using both point sampling (instantaneous measures of swim position at first surface after each 2.5-min interval) and PAS (calculated from the continuous data on observed onsets and terminations of calf position swimming). I compared the percent time swimming in calf position for 19 calves in my longitudinal sample. Each calf was observed during focal-individual follows for 10–15 h per yr for one or more years. The sum of observation hours for all calves across all years was 750. Each infant had a PAS and point sampling percentage for each year, for a total of 40 calf-years (an average of two years or samples per calf). The PAS percentages and point sampling percentages were correlated by treating each infant observation year as independent. This yielded a significant value (Spearman's *rho* = 0.97, $n = 40$, $P < 0.0001$). The mean percent error rate for PAS

was 3.2%. Several methodologists (*e.g.,* Tyler 1979) have demonstrated the validity of PAS and point sampling using statistical models, but no one has contrasted these methods using actual observational data.

Incident sampling—Alternate names: all-event sampling, all-occurence sampling, all-animals sampling. Incident sampling entails scoring all behavioral events of a specific type in a group. This method is not applicable for most behavioral states. The observability of the events is key to the success of this method. The behaviors themselves must be obvious enough to alert the observer (*e.g.,* breaching). In addition, the observer must be able to record all the events regardless of how many animals are present. Thus, for incident sampling to be successful, the behavior must be sufficiently infrequent to allow complete recording for the group. As Altmann points out, this is a form of continuous sampling (referred to as focal animal sampling in her paper) of a group for a restricted set of behaviors. However, with continuous sampling, the individual animals are identified. During incident sampling, distinctions between subjects may or may not be made. Observers should still make every effort to determine whether the same or different animals are exhibiting the behavior(s). Two areas can be problematic. First, animals may tend to repeat displays, thus becoming overrepresented in the dataset. Second, if group composition changes often (Table 1) it may be difficult to calculate event rate as a function of group size or structure.

In the survey, 16% of the studies used incident sampling. Studies that identified their method as "focal-group-sampling" were sometimes coded as using incident sampling *if* they scored a very restricted set of obvious behaviors. For example, breaches and lobtails are so visible that observers are likely to detect each occurrence within a group. Similarly, if observers could determine how many animals were within acoustic range, then incident sampling could apply to studies of vocalizations.

Incident or all-event sampling is valuable for cetacean researchers who wish to focus on specific dramatic or easily recognized events that involve few animals. For example, observers could use this method to compare successful and unsuccessful killer whale hunting attempts *via* beaching. Dramatic surface percussive displays, such as breaching or lobtailing, can also be observed reliably in groups (Waters and Whitehead 1990). Incident sampling requires that the observer systematically record every event. Incident sampling may be adopted with a group protocol, but the observer should still be able to distinguish which animal (or age/sex class member) is engaging in the behavior to avoid misattributing the behavior pattern equally to all individuals.

Sequence sampling—In sequence sampling, the observer focuses on sequences of behavior or on particular interactions, rather than individuals, and systematically records all relevant behaviors that occur during the event(s), maintaining the sequences of behaviors in the record (Altmann 1974). Parameters defining how an interaction begins and terminates must be specified in sequence sampling. For example, observers may score sequences of behaviors among surface-active groups of humpback whales by defining the beginning of the sequence as "male humpback whale challenges principal escort by ap-

proaching within 50 m" and terminating the sequence when either the chal-
lenger or the principal escort separates more than 50 m from the female for
30 min. It is distinct from incident sampling because during sequence sam-
pling multiple events may occur in a group, but the observer focuses on the
start of the first event seen and records that particular event or interaction to
completion (even if other group members engage in similar behaviors at the
same time). Only one of the studies surveyed used sequence sampling.

Sequence sampling is recommended for very observable behaviors but can
be applied under a broad range of field conditions (see Table 1). The researcher
must be able to determine when the sequence begins and ends and be able to
discriminate between, although not necessarily identify, the interactants. The
individual animals may be hard to identify if they communicate over long
ranges and move in and out of the observer's view. For example, if the re-
searcher wants to know whether breaching animals attract or repel others, s/
he can use sequence sampling to test whether animals are likely to approach
or leave the animal who breached (*e.g.*, within 10 min of the first breach).
Sequence sampling is excellent for determining the conditional probabilities
of behavior sequences.

RECOMMENDATIONS

Although tightly focused and informative studies peppered the literature,
the survey of CJZ and MMS revealed two shortcomings in studies of cetacean
behavior. First, a large proportion of studies used *ad libitum* sampling, which
is replete with bias. Second, there was little consistency in reporting behavioral
data, sampling methods, hours of observation, behavioral and group defini-
tions, and number of subjects. This information is critical to evaluation of the
claims of any study. Furthermore, the comparability of results and hopes of
replication are lost if this information is not provided. Information about
equipment (cameras, film, boat, motor, tape recorders, *etc.*) was frequently
reported, but basic methodological information regarding subjects and pro-
tocol was sometimes lacking.

To allow evaluation of statistical power, it is important to indicate the
number of animals observed and how long each animal was observed. Re-
searchers usually reported the number of animals identified in the population
but not the number of animals observed. Sample sizes can be determined only
when animals are observed individually. Similarly, only a few studies treated
each subject, rather than each observation, as independent. As Milinski (1997)
points out, this is a form of pseudoreplication, one of the "deadly sins" in the
study of behavior.

Group size and definition of group are needed, as well as a statement of
protocol used if group composition changed. I recommend proximity-based
measures, because this method is quantifiable and does not rely on behavioral
sampling to determine group membership. "Coordinated-behavior" definitions
make implicit assumptions about proximity, because observers cannot assess
the activities of animals who are kilometers away. Furthermore, individuals at

close range may be engaged in more than one activity that is shared with those farther away. An explicit definition of group, and a rationale for the definition in terms of the study goals, is essential.

Specific details of how behaviors were defined and how behaviors were sampled should be included, using either standardized terms from the ethological literature or descriptions. Only a few studies of those reviewed used standard sampling terms described by Altmann (1974) and others. All studies of cetacean behavior should provide the basic methodological information described in this review. Following are more specific recommendations regarding the uses of different sampling protocols and techniques, with special consideration of how to reconcile unbiased observations under ideal conditions with the imperfect conditions that cetacean biologists face. Additional detailed discussion of methods and statistical comparisons of different observational methods are available elsewhere (see Altmann 1974, Dunbar 1976, Tyler 1979, Martin and Bateson 1986, Mann *et al.* 1991, Rogosa and Ghandour 1991).

Follow Protocol

Sampling biases in group-follows and surveys can be minimized by using scan sampling, incident sampling, or sequence sampling methods. Monitoring groups (surveys and group follows) is valuable when the goal is to gain a "snapshot" of group life through surveys (*e.g.,* habitat use by foraging compared to resting groups) when observation conditions do not permit tracking of individuals (but tracking of groups is possible), when research questions focus on the synchrony or coordination of group members, or when the study focuses on sequences of very observable behaviors. To identify the predominant group activity during a survey, observers can scan-sample >50% of the group members. This may be possible for large groups of dolphins or whales that are visible at the surface for long enough to be scanned. If group sizes are in the hundreds or thousands, then a protocol for random scan sampling of a subset of the group may be developed. Sampling biases, such as number of scans per individual as a function of group size, must be taken into account during data reduction. Researchers may track a group and record all occurrences of certain very obvious behaviors, thus using incident sampling. Sequence sampling may also be applied if observers are interested in sequences of behavior during specific types of interactions, such as cooperative hunting in killer whales. Focal-group sampling is not recommended, because this method does not explicitly or systematically sample individuals or behaviors in groups.

Focal-individual sampling methods with the application of continuous, point, and/or predominant-activity sampling, are critical tools for insuring reliable estimates of behavior. The individual is the natural unit of analysis for behavior. Data can be pooled by individual, and data from different individuals can often be treated as independent (Machlis *et al.* 1985). When observers cannot identify or discriminate which individual produces a behavior (*e.g.,* during scan- or all-event sampling), this makes it difficult to identify

independent units for statistical analysis. This can present problems with the group-follow protocol.

General Sampling Problems: Deep Divers, Large Groups, and Correspondence Between Surface and Subsurface Behavior

The correspondence between surface (observable) and subsurface (often unobservable) behavior is unknown for most studies of cetaceans, but such information could help in assessing the biases in relying on surface behavior alone. For example, some animals forage at depth and not during surfacing bouts. One option is to define the surfacing breaks between foraging dives as part of the continuous state of "foraging." Other behaviors may not change at the surface (*e.g.,* socializing, resting, traveling). Although it is still valuable to record surface behavior even though no correspondence to subsurface behavior is implied, such sampling decisions should be indicated clearly in the methods section of the article.

All ethologists face difficult decisions when their subjects disappear from view. These "out-of-sight" periods may be treated differently, depending on the animal under study. Nesting or burrowing animals may be most likely to disappear from view when feeding their young or engaging in sexual behavior. Some cetaceans may disappear for long periods when foraging. Deleting the time that animals are "out of sight" is desirable when animal visibility is not determined by or biased by animal activity. Other strategies for using "out-of-sight" periods are needed for cetaceans, because some of their activities co-vary with diving periods (when animals are most likely to be out of view).

Cetaceans may rest, travel, socialize, or engage in other diverse behaviors at or near the surface. Certain types of social behavior may occur at depth, but certainly not all of them do. If the out-of-sight periods tend to be short relative to the bout lengths of the behavior, then ethologists might designate the activity for the out-of-sight period as either the last activity seen before the subject disappeared or the first activity seen when the animal reappears. Alternatively, the observer may consider a state continuous only if the last activity seen and the first activity seen after a brief out-of-sight period are the same. This strategy is not effective if some behaviors, such as foraging or hunting, consistently occur out of an observer's view. In such circumstances, cetacean observers may detect foraging using other cues, including bubble patterns at the surface, echolocation clicks, or intermittent sightings of fish catches or chases. Sometimes there are signs at surfacing of what the animal was doing on the preceding dive. For example, bottom-feeding whales may surface with mud streaming from the mouth (Würsig *et al.* 1985). Thus, observers may use fleeting observations of events to define states or indirect cues (rapid surfacing, changing direction) to define "unobservable" states. One may also tag an individual to track behaviors at depth for comparison with cues visible at the surface. By combining these techniques, observers can develop behavioral definitions and systematic protocols to capture the range of cetacean behaviors.

Sampling from large groups of hundreds or thousands (*e.g.*, Scott and Perryman 1991) can prove to be difficult. However, Östman (1994) has successfully completed short focal follows on *Stenella longirostris* in Hawaii, suggesting that it is possible to stay with individuals long enough to get a 5-min sample using an underwater observation booth built into a vessel. Scan sampling of large groups of unidentified animals is another possible approach. In such circumstances, researchers would need to scan a subset, possibly by randomly selecting subgroups or particular age/sex classes. Videotaping for later scoring or photogrammetric analysis (*e.g.*, Scott and Perryman 1991) can provide data on association and behavior.

Although observation conditions for studying cetaceans are rarely ideal, proximity between animals, synchrony in surfacing, and basic activities can be recorded systematically under most sampling conditions. Cetacean behavioral research could benefit greatly from a wider use of quantitative sampling techniques. If applied carefully, such partial but systematic observations can help elucidate cetacean social and behavioral ecology.

LITERATURE CITED

ALTMANN, J. 1974. Observational study of behavior: Sampling methods. Behaviour 49: 227–267.
CLARK, C. W. 1980. A real-time direction finding device for determining the bearing to the underwater sounds of southern right whales, (*Eubalaena australis*). Journal of the Acoustical Society of America 68:508–511.
DUNBAR, R. I. M. 1976. Some aspects of research design and their implications in the observational study of behaviour. Behaviour 58:79–98.
FRAGASZY, D. M., S. BOINSKI AND J. WHIPPLE. 1992. Behavioral sampling in the field: Comparison of individual and group sampling methods. American Journal of Primatology 26:259–275.
FRANKEL, A. S., C.W. CLARK, L.M. HERMAN AND C.M. GABRIELE. 1995. Spatial distribution, habitat utilization, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawaii determined using acoustic and visual techniques. Canadian Journal of Zoology 73:1134–1146.
FREITAG, L., AND P. L. TYACK. 1993. Passive acoustic localization of Atlantic bottlenose dolphin whistles and clicks. Journal of the Acoustical Society of America 93: 2197–2205.
HAMMOND, P. S. , S. A. MIZROCH AND G.P. DONOVAN. EDS. 1990. Individual recognition of cetaceans: Use of photo-identification and other techniques to estimate popu-

lation parameters. Report of the International Whaling Commission (Special Issue 12).

HUTT, S. J., AND C. HUTT. 1970. Direct observation and measurement of behavior. Charles C. Thomas, Springfield, IL.

MACHLIS, L., P. W. D. DODD AND J. C. FENTRESS. 1985. The pooling fallacy: Problems arising when individuals contribute more than one observation to the data set. Zeitschrift für Tierpsychologie 68:201–214.

MANN, J., AND B. SMUTS. In press. Behavioural development of wild bottlenose dolphin newborns. Behaviour.

MANN, J., T. TEN HAVE, J. W. PLUNKETT AND S. J. MEISELS. 1991. Time sampling: A methodological critique. Child Development 62:227–241.

MARTIN, P., AND P. BATESON. 1986. Measuring behaviour: An introductory guide. Cambridge University Press, Cambridge, UK.

MILINSKI, M. 1997. How to avoid seven deadly sins in the study of behavior. Advances in the Study of Behavior 26:160–180.

ÖSTMAN, J. S. O. 1994. Social organization and social behavior of Hawai'ian spinner dolphins (Stenella longiristris). Ph.D. dissertation, The University of California, Santa Cruz, CA. 114 pp.

ROGOSA, D., AND G. GHANDOUR. 1991. Statistical models for behavioral observations. Journal of Educational Statistics 16:157–252.

SAMUELS, A., AND T. GIFFORD. 1997. A quantitative assessment of dominance relations among bottlenose dolphins. Marine Mammal Science 13:70–99.

SCOTT, M. D., AND W. L. PERRYMAN. 1991. Using aerial photogrammetry to study dolphin school structure. Pages 227–241 in K. Pryor and K. S. Norris, eds. Dolphin societies: Discoveries and puzzles. University of California Press, Berkeley, CA.

SHANE, S. H. 1990. Behavior and ecology of the bottlenose dolphin at Sanibel Island, Florida. Pages 245–265 in S. Leatherwood and R. R. Reeves, eds. The bottlenose dolphin. Academic Press, Inc., New York, NY.

TYACK, P. 1985. An optical telemetry device to identify which dolphin produces a sound. Journal of the Acoustical Society of America 78:1892–1895.

TYACK, P. L., AND C. A. RECCHIA. 1991. A datalogger to identify vocalizing dolphins. Journal of the Acoustical Society of America 90:1668–1671.

TYLER, S. 1979. Time-sampling: A matter of convention. Animal Behaviour 27:801–810.

WATERS, S., AND H. WHITEHEAD. 1990. Aerial behaviour in sperm whales. Canadian Journal of Zoology 68:2076–2082.

WHITEHEAD, H. 1995. Investigating structure and temporal scale in social organizations using identified individuals. Behavioral Ecology 6:199–208.

WHITEHEAD, H. 1997. Analyzing animal social structure. Animal Behaviour 53:1053–1067.

WHITEHEAD, H., AND L. WEILGART. 1991. Patterns of visually observable behaviour and vocalizations in groups of female sperm whales. Behaviour 118:275–296.

WILLIAMS, G. C. 1966. Adaptation and natural selection. Princeton University Press, Princeton, NJ.

WÜRSIG, B., E. M. DORSEY, M. A. FRAKER, R. S. PAYNE AND W. J. RICHARDSON. 1985. Behavior of bowhead whales, Balaena mysticetus, summering in the Beaufort Sea: A description. Fishery Bulletin, U.S. 83:357–377.

APPENDIX 1

I conducted a search of the Science Citation Index (Institute for Scientific Information, Inc., 1996) using all 39 cetacean genera for keywords. The search yielded 846 studies. To determine whether the study focused on cetacean behavior in the wild, each title was examined for keywords: behavior, acoustics, observations, field, captive, wild, and a description of the geographic location. The 846 studies were classified according to 14 subjects (e.g., molecular, neurophysiological, conservation, behavior); many studies received more than one classification. Of the total, 125 studies (14.8%) in 28 journals focused on cetacean behavior ($n = 106$) or vocalizations ($n = 19$) in the wild. Studies that appeared to focus strictly on population assessment (e.g., aerial counts) were not classified as behavioral studies. CJZ and MMS published 60% of the studies. The next most popular journal for cetacean behavior studies was the *Journal of Mammalogy*, which published only 5% ($n = 6$). Thus, CJZ and MMS clearly represent the majority of relevant peer-reviewed publications in the Science Citation Index (SCI) and were chosen for in-depth analysis for this reason. Based on a comparison of Impact Factors (Impact Factor = the average number of times a paper in a journal is cited in the Science Citation Index within two years of publication), MMS and CJZ represent mid-level quality in research. Approximately an equal number of cetacean studies ranked above ($n = 21$) and below ($n = 22$) MMS and CJZ, with six studies (published in the *Journal of Mammalogy*) ranking the same as CJZ and one study ranking below CJZ but above MMS.

Because the SCI search may have missed studies that did not use genera for keywords, I surveyed all issues of CJZ and MMS for all studies of free-ranging cetacean behavior and vocalizations published during 1989–1995. The SCI search by genera did not miss any studies, but one CJZ study I classified as a "behavior study" from the title was excluded from the review because the article's content did not include behavior or vocalizations.

The basic unit of analysis was one published article. Of 144 authors, 21 (14.6%) published more than one study. Of these, 14 published two studies, six published between three and five studies and one researcher authored or co-authored nine studies (using five different sampling methods).

Species studied—Twenty-four cetacean species were studied in the 74 papers I reviewed. Studies of humpback whales (*Megaptera novaeangliae*) were the most common (24%), followed by those of sperm whales (*Physeter macrocephalus*, 16%), and killer whales (*Orcinus orca*, 16%). Bottlenose dolphins (*Tursiops truncatus*) were represented in 9% of the studies. Bowhead (*Balaena mysticetus*) and gray whales (*Eschrictius robustus*) were each represented in 5% of the studies. Narwhals (*Monodon monoceros*), Bryde's whales (*Balaenoptera edeni*) and belugas (*Delphinapterus leucas*) were each represented in 4% of the studies. Seven studies included more than one species. Of the 15 remaining species, only one or two studies were published on each.

Age and sex classes sampled—Twenty studies focused on animals of a particular age or sex class. Otherwise, observers either did not know the ages or sexes of animals, or they observed animals of all ages and both sexes. No studies made quantitative comparisons between age or sex classes.

Number of animals sampled—The range of animals sampled was one to >1,000. Typically, researchers reported the number of animals that they surveyed or identified but not how many animals were observed or sampled. In 45% of the studies, the number of animals identified and/or sampled could not be determined. When groups of animals were sampled (e.g., in surveys or group-follows), 59% of the time ($n = 44$ studies) I could not distinguish between how many animals were individually identified and how many were actually observed.

Individual identification—In 66% ($n = 49$) of the studies, researchers could identify individual animals. In 34% ($n = 25$), individual animals were not identified, although in about half of those (48%, $n = 12$), physical features of the species and reference to

other studies indicated that individual identification was possible. In all studies that could identify individuals, rates of individual behavior or vocalizations were not reported unless the study involved only one or two animals ($n = 3$ studies).

Observation/recording hours—Twenty-eight studies (38%) did not report how much time was spent observing or recording behavior. Some of these studies reported the hours spent at sea, but this information is only useful (as an indicator of sighting effort) when observation hours are also presented. Sixty-two percent of the studies reported the number of surveys conducted or the number of hours spent observing animals. Only four studies treated each subject, rather than each observation, as independent for behavioral analysis.

Group definitions and group size—Definition of "group" is essential to assess the validity of behavioral sampling from groups. I indicated whether or not "group" was defined, and if it was defined, I noted the definition. Group-size definition was not relevant in nineteen studies (*e.g.,* only one animal was studied). For 55 studies (surveys, group-follows, anecdotes), groups of animals were observed. In 42 of these (76%), researchers did not define "group." If group was defined ($n = 13$), proximity-based measures (*e.g.,* within 10 m) or coordinated behavior (all visible animals engaged in the same activity) were the two most common definitions of group.

Group-size means or ranges (minimum and maximum group size) were reported in 38 studies. Seventeen studies did not report group size information.

Types of behavior and vocalizations recorded—Behaviors recorded in each study were classified as breathing (14%), diving (30%), resting (24%), surface displays (8%), socializing (30%), feeding or foraging (43%), traveling (27%), or "other." Association (proportion of time animals together) was recorded in 22% of the studies and vocalizations were recorded in 38%. "Other" included births (1%) or other unusual events (16%). Over half of the studies (57%) provided definitions for one or more of the behaviors recorded. Studies were classified according to the behaviors that researchers recorded, not which data were analyzed. If an ethogram was given, this was noted. No studies reported measures of inter-observer or intra-observer reliability.

Sound recording was used either to find or track whales or to describe the vocal repertoire of a species, group, or individual. Of the 29 papers that recorded dolphin or whale vocalizations, 28% identified species-specific repertoires or group-specific repertoires. In one study, species-typical vocalizations were monitored to test for the presence of the species. Identification of individual repertoires within groups was limited to anecdotes, but one study used passive acoustic localization to identify which animal was likely to be vocalizing. Fourteen percent of the studies that recorded cetacean vocalizations calculated vocalization rates for groups of animals but not for individuals. Solitary individuals were acoustically recorded for portions of one study.

All 74 studies were classified according to sampling method: *ad libitum, continuous, focal-group, one-zero, point, scan, predominant activity, sequence,* and *incident sampling.* If a study used more than one sampling method (six studies), both were scored. The results of these classifications are reported in the main text.