

Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales

Patrick J. O. Miller,^{a)} Ricardo N. Antunes, Paul J. Wensveen, Filipa I. P. Samarra, Ana Catarina Alves, and Peter L. Tyack
Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife KY16 8LB, Scotland

Petter H. Kvadsheim and Lars Kleivane
Maritime Systems Division, Norwegian Defence Research Establishment (FFI), NO-3191 Horten, Norway

Frans-Peter A. Lam and Michael A. Ainslie
Acoustics and Sonar, Netherlands Organisation for Applied Scientific Research (TNO), P.O. Box 96864 The Hague, 2509 JG, The Netherlands

Len Thomas
Centre for Research into Ecological and Environmental Modelling, University of St Andrews, Fife KY16 9QQ, Scotland

(Received 5 October 2012; revised 8 November 2013; accepted 17 December 2013)

Eight experimentally controlled exposures to 1–2 kHz or 6–7 kHz sonar signals were conducted with four killer whale groups. The source level and proximity of the source were increased during each exposure in order to reveal response thresholds. Detailed inspection of movements during each exposure session revealed sustained changes in speed and travel direction judged to be avoidance responses during six of eight sessions. Following methods developed for Phase-I clinical trials in human medicine, response thresholds ranging from 94 to 164 dB re 1 μ Pa received sound pressure level (SPL) were fitted to Bayesian dose-response functions. Thresholds did not consistently differ by sonar frequency or whether a group had previously been exposed, with a mean SPL response threshold of 142 ± 15 dB (mean \pm s.d.). High levels of between- and within-individual variability were identified, indicating that thresholds depended upon other undefined contextual variables. The dose-response functions indicate that some killer whales started to avoid sonar at received SPL below thresholds assumed by the U.S. Navy. The predicted extent of habitat over which avoidance reactions occur depends upon whether whales responded to proximity or received SPL of the sonar or both, but was large enough to raise concerns about biological consequences to the whales. © 2014 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4861346>]

PACS number(s): 43.80.Nd [JJF]

Pages: 975–993

I. INTRODUCTION

Observations of effects of sonar on cetaceans are relatively limited (Nowacek *et al.*, 2007), but many studies indicate that avoidance of anthropogenic sound sources is a common response. Migrating animals show avoidance by changes in the travel path around a noise source (Richardson *et al.*, 1995; McCauley *et al.*, 2000; Buck and Tyack, 2000). For non-migrating animals, avoidance may involve a switch from ongoing activities to traveling away from the sound (Nowacek *et al.*, 2004; Lusseau *et al.*, 2009; Goldbogen *et al.*, 2013). Avoidance reactions can potentially lead to negative effects including habitat loss via long-term avoidance of noisy locations (Morton and Symonds, 2002), energetic costs of increased locomotion and decreased feeding time (Lusseau *et al.*, 2009; Goldbogen *et al.*, 2013), and atypical mass-stranding events dominated by beaked whales

(D'Amico *et al.*, 2009) which could be the consequence of strong avoidance reactions to the sonar (Cox *et al.*, 2006; Tyack *et al.*, 2011; DeRuiter *et al.*, 2013). In addition to negative effects on cetaceans themselves, reactions to human activities may affect our own ability to benefit from them. There was strong public concern when whale watchers and researchers observed killer whales being unintentionally exposed to military sonar transmissions within Haro Strait along the U.S. and Canadian border (NMFS, 2005).

Whale-watching is a globally important economic activity (Cisneros-Montemayor *et al.*, 2010) which would be negatively affected by any decline in the numbers of animals available for observation. Multi-national naval exercises in Vestfjorden, Norway, in 2000 (WWF-Norway, 2001) and 2006 were blamed for reduced numbers of killer whales due to the use of active high power anti-submarine sonar (see Sec. IV).

The U.S. Navy uses a dose-response curve to estimate the cumulative probability that a cetacean might be “harassed,” under the definition of the U.S. Marine Mammal Protection Act, as a function of the maximum received SPL of the sonar

^{a)}Author to whom correspondence should be addressed. Electronic mail: pm29@st-andrews.ac.uk

(U.S. Navy, 2008, 2012). However, accurate assessment of the potential environmental impact of sonar activities is hindered by a lack of targeted dose-response studies to establish thresholds at which free-ranging whales react to sonar sources. It is therefore valuable to report the acoustic received level of a stimulus associated with animal responses (Southall *et al.*, 2007; Houser *et al.*, 2013). Received levels can be measured at or near the whale, and/or can be estimated using acoustic propagation models for well characterized sound sources and environments. Metrics for acoustic received level that have been used include SPL, cumulated sound exposure level (SEL), and SPL or SEL weighted by a function related to the hearing curve such as sensation level, A-weighting for humans (Kinsler *et al.*, 1982), and M-weighting for marine mammals (Southall *et al.*, 2007). SPL is the received level “dose” most often reported in noise disturbance studies. However, it is unknown which characteristics of a received acoustic signal best predict the likelihood of behavioral responses, suggesting the utility of reporting received levels using several different acoustic metrics (Madsen *et al.*, 2006). In addition to the acoustic features of the received sonar signal, the proximity, speed, and direction of motion of the source with respect to the whale may also influence the likelihood or thresholds of behavioral response. Finally, other factors such as behavioral context, food availability, exposure history, and individual differences may modify the acoustic levels to which animals respond to noise (Southall *et al.*, 2007; Ellison *et al.*, 2012).

These issues stimulated the study reported here, which was designed to quantify dose-response relationships of cetaceans responding to sonar using carefully conducted dose-escalation experiments, and methodologies developed for phase-I clinical trials in human medicine (Simon *et al.*, 1997). Phase-I clinical trials are typically the first trials undertaken with human patients in the evaluation of new drug treatments, and focus on typifying the dose-response relationship sufficiently (often with small sample sizes) in order to identify a safe dose for further trials of efficacy and risk. In these trials, patient responses are scored by a physician on a case-by-case basis. Advanced methods in design and analysis of phase-I clinical trial data use underlying dose-response models to continually update the underlying dose-response relationship as data become available (O’Quigley *et al.*, 1990). Bayesian models can be particularly useful to build preliminary dose-response functions in the face of limited data (O’Quigley and Conaway, 2010), are useful to account for uncertainty in the observational data, and can incorporate any prior information about response parameters that might be available.

In the present work, we are concerned with the risk of behavioral avoidance responses (a type of harassment) of whales by sonar at sea, and though there is good evidence that whales sometimes avoid powerful sound sources, there has been little information to build dose-response relationships for whales at sea. Some of the data used for the currently applied U.S.-Navy curve came from captive delphinids that were positively reinforced for tolerating exposure to tonal sounds used to test for temporary shifts in hearing thresholds (HTs) (U.S. Navy, 2008), a very different situation from sonar exposure in the wild. One data point of 169 dB re 1 μ Pa received SPL came from a reconstruction of killer whale reactions to a naval

sonar exercise (NMFS, 2005). However, that received level data point was taken as the maximum exposure at the closest point of approach, not the level at which the animals began to exhibit behavioral reactions. Another point used in the development of the U.S. Navy curve was a study on responses of North Atlantic right whales to alarm signals (Nowacek *et al.*, 2004), which are quite different from naval sonar signals. Given extremely limited information about the thresholds at which cetaceans in the wild respond to sonar, there is a clear benefit to updating the dose-response functions as data become available from carefully conducted experiments with whales at sea. Thus, the goal of this paper is to use experimental data from killer whales exposed to sonar at sea in a fashion analogous to Phase-I trials in humans.

The study used two sonar frequency bands (1–2 kHz and 6–7 kHz) in order to explore the influence of sonar frequency, in relation to frequency-dependent hearing sensitivity, on the thresholds of behavioral response. The audiogram of killer whales (Szymanski *et al.*, 1999) suggests that killer whale hearing is 20–30 dB more sensitive at 6–7 kHz than at 1–2 kHz. Sensation level is defined with respect to auditory sensitivity, so the sensation level of a 6–7 kHz signal would be roughly 20–30 dB higher than that of a 1–2 kHz signal at the same SPL. If the difference in hearing sensitivity of the two sonar frequencies is an important predictor of behavioral response, then we expect differences in the received SPL associated with response thresholds for the two different sonar frequencies. In such a case, representing the received sonar signal as sensation level could be an effective means to account for the influence of hearing sensitivity in that frequency band on the likelihood of behavioral response.

Avoidance was chosen after the field work as a suitable response parameter for this first attempt to develop dose-response relationships. Avoidance is an important type of behavioral response, which is commonly documented and has an additional benefit that it can be quite easily characterized and identified in behavioral records as movement away from a source, or from the path of a moving source. The killer whale is an appropriate study species because killer whales have previously been reported to respond to naval sonar transmissions (WWF-Norway, 2001), and some hearing curves are known (Hall and Johnson, 1972; Szymanski *et al.*, 1999). However, they are difficult to work with in field experimental designs, and we expected the sample of experiments we might obtain from at-sea experiments to be few in number. Therefore, during each exposure session, we gradually increased or “escalated” the sonar dose received by the whale in order to maximize our ability to identify a reaction threshold for each exposure session. Detailed inspection of data from each exposed whale was undertaken on a case-by-case basis to identify whether an avoidance reaction took place, and the precise time when the reaction started to be apparent in the data record (Miller *et al.*, 2012). Different sonar dose measures at the time that avoidance started were considered to be the response threshold for that subject whale for that session, and the observed thresholds were used to estimate parameters in Bayesian dose-response functions. We also consider how to extrapolate from our dose-response results to predict responses to naval sonar activities.

II. METHODS

A. Experimental procedure

We conducted experiments exposing free-ranging killer whales to sonar in 2006, 2008, and 2009, and details of each experiment are available in a technical report (Miller *et al.*, 2011) and descriptions of behavioral changes provided in Miller *et al.* (2012). Suction cup tags (Dtag; Johnson and Tyack, 2003) that record sound, depth, three-axis accelerometer, and three-axis magnetometer data were attached to six individuals in four different groups, and subjects were tracked from an observation vessel. The behavior and location of the subject's group was monitored from a 5 m length workboat in 2006 and from the 29 m MS *Strønstad* in 2008 and 2009. A mitigation protocol was in place to cease transmission in case any reaction appeared to entail a risk of harm to any study subject.

Following a pre-exposure period, the subject was approached by the source vessel RV *HU Sverdrup II* as it transmitted sonar pulses of 1 s duration every 20 s (Fig. 1). Throughout each exposure session, all sonar transmissions were frequency modulated hyperbolic (Ainslie, 2010) upsweeps (except for the final exposure session which used a hyperbolic downsweep) within one of the two different frequency bands. For the 1–2 kHz and 6–7 kHz sonar bands, respectively, the source level (Morfey, 2001) started at 152

and 156 dB re $1 \mu\text{Pa m}$ (150 and 138 dB in 2006), and was gradually increased (“ramped-up”) over 10 min to 214 and 199 dB (209 and 197 dB in 2006). The source vessel adjusted course to approach the tagged whale until 1 km distant, at which point the course of the source vessel was fixed. Transmissions ceased 5 min after passing the closest point of approach to the whale, or if any condition potentially harmful to the study animals was observed (such as calf separation). In 2006, only one exposure session was conducted with each of two subject groups. Given the difficulty of accomplishing these experiments with free-ranging animals at sea, we attempted to gain more information by conducting multiple exposures on tagged subjects in 2008 and 2009, with a gap of at least 1 h between exposures (Table I). Exposure sessions are labeled as “experiment number–exposure number within group” (e.g., “2-1” means the first exposure session within experiment number 2).

Sightings were taken from the observation vessel every 2–5 min to track the tagged whale(s). The position of the whale surfacing was fixed using the range and bearing to the animal from the observation vessel, whose GPS position was recorded every 10 s. Distance to the whale was measured by laser range-finders whenever possible, or estimated by eye, and bearing to the whale was measured using a protractor mounted to the observation vessel combined with the ship's

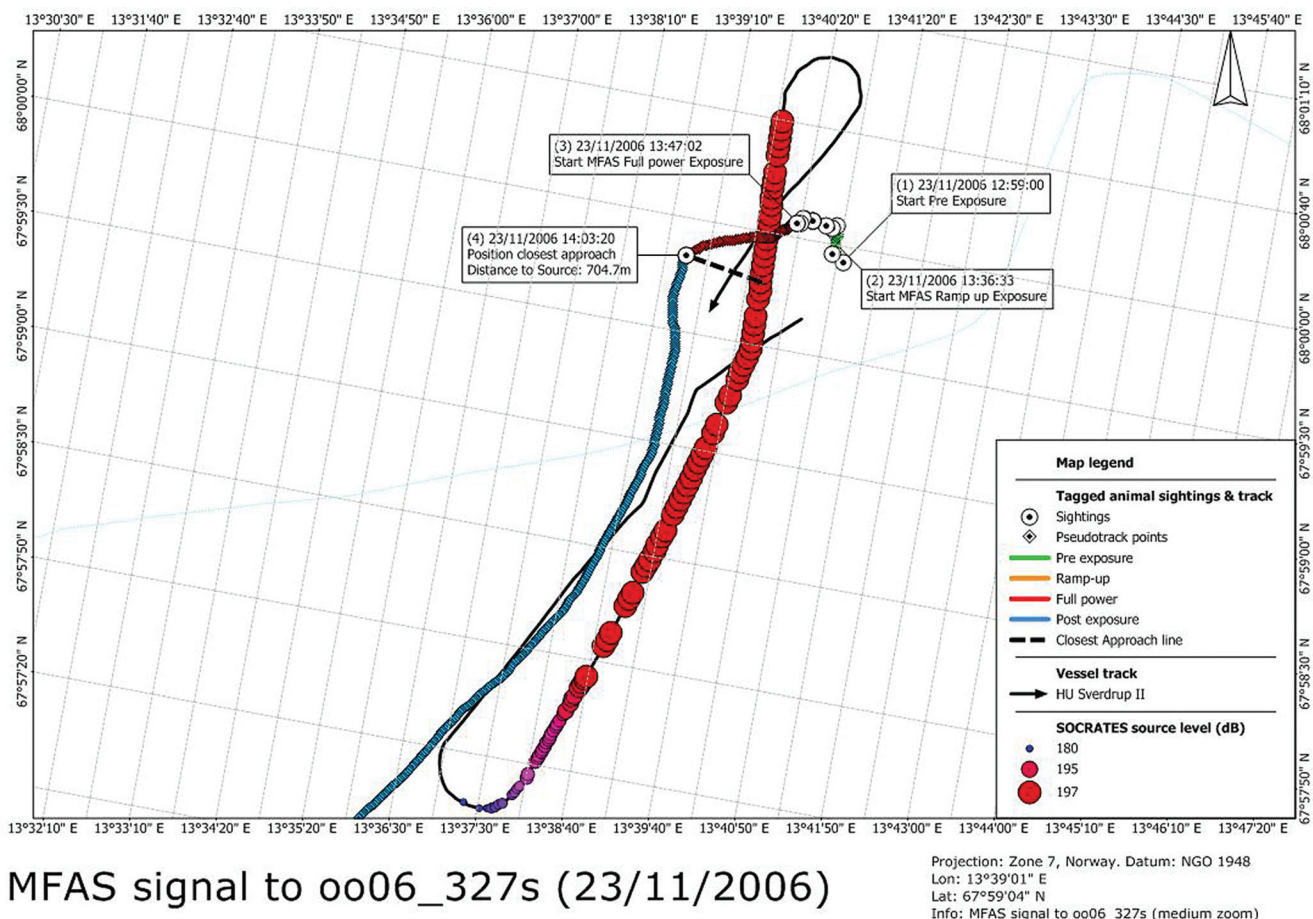


FIG. 1. Geometry of dose-escalation experiment 2-1. The source vessel track is shown in the thinner black line with colored red dots representing the position of the Socrates sonar source and the source levels of transmissions, up to full-power level of 197 dB re $1 \mu\text{Pa m}$ in this example. The track of the whale is indicated with the thick colored line, with sightings indicated with large circles and dead-reckoned track points (“pseudotrack”) indicated with small diamonds. Pre-exposure movement is shown in green, exposure in red, and post-exposure in blue.

TABLE I. Evaluation of objective criteria to determine response for each exposure session. Horizontal and vertical arrows refer to movement away from the source path (horizontal arrows) and source location (vertical arrows). In some cases both were observed within a session.

	oo317s	oo327s(t)	oo149a	oo149a	oo149a	oo144a(b)	oo144a(b)	oo144a(b)
Tag id(s):	13/11/06	23/11/06	28/05/08	28/05/08	28/05/08	24/05/08	24/05/08	24/05/08
Date Tag-on time:	13:32:11	12:42:13	09:01:39	09:01:39	09:01:39	09:58:53	09:58:53	09:58:53
Session	1-1	2-1	3-1	3-2	3-3	4-1	4-2	4-3
Time start	14:10:00	13:36:00	12:48:00	14:56:00	22:38:00	14:13:00	16:15:00	21:13:00
End (UTC) sonar band	14:43:01	14:10:01	13:40:41	15:46:01	23:08:21	14:47:01	17:14:01	21:51:01
	1-2 kHz	6-7 kHz	6-7 kHz	1-2 kHz	6-7 kHz	1-2 kHz	6-7 kHz	1-2 kHz ^a
Avoidance signature?	Y →	Y →↑	N	N	Y → brief	Y →	Y →↑	Y →↑
Change in direction?	Y, gradual	Y	N	N	Y, brief	Y	Y	Y
↑ linearity?	N	Y	N	N	N	Y	N	N
↑ speed?	Y	Y	N	N	N	Y	Y, variable	Y
Break-point statistic	0.98	<0.001	0.90	0.91	1.0	<0.001	0.299	<0.001
Other behavioral indicators		stopped feeding			separation of calf, vocal	stopped feeding, vocal, group		
Other relevant factors?	short base-line	short base-line	in shallow water	in shallow water	narrow fjord			
Final Conclusion	Y	Y	N	N	Y	Y	Y	Y
Confidence	med	high	high	high	med	high	high	high
Onset time	14:31:30	13:56:25	none ^b	none ^b	22:41:49	14:13:30	16:17:40	21:33:10
Distance (km)	4.6	2.4	<i>1.5</i>	<i>1.2</i>	0.7	7.8	8.9	3.2
SPL (dB re 1 μPa)	150	138	<i>142</i>	<i>166</i>	133	94	94	164
Sensation level (dB re 1)	61	83	<i>87</i>	<i>81</i>	78	3	39	83
SEL (dB re 1 μPa ² s)	158	143	<i>149</i>	<i>176</i>	133	94	95	171

^aIn this session, the waveform was a hyperbolic downsweep.

^bFor these exposure sessions, the maximum acoustic values, and minimum approach distance are reported in italic type face.

heading estimated from a digital compass and GPS course over ground. The observation vessel sought to maintain a distance of 100–200 m from the tagged whale throughout the follow. Systematic sightings were taken of one focal whale, chosen just after tagging based upon the quality of the tag placement, while any other tagged whales were only sighted opportunistically. Thus, the movement track was of one particular tagged whale within the group.

Upon recovery of the tags, pressure data were converted to depth using calibrated values, compensating for temperature effects. Similarly the accelerometer and magnetometer output was converted to field strength on each axis (Johnson and Tyack, 2003), and the pitch, roll, and heading of the whale were calculated following published methods (Johnson and Tyack, 2003; Miller *et al.*, 2004). A constant speed dead-reckoned track (Miller *et al.*, 2009) was first calculated for the periods between consecutive sightings. Dead-reckoning started at the first sighting and used an average speed calculated by the ratio of the distance and the difference in time between sightings. The dead-reckoned track points were then corrected by adding a two-dimensional x-y vector whose magnitude and angle would make the point corresponding to the time of the next surfacing match the sighting position at that surfacing. Points leading up to the surfacing were adjusted by interpolating the magnitude of the correction vector linearly against time, from zero correction at the previous surfacing to the full correction vector for the next surfacing and thereafter. This process was repeated sequentially from the first to the last sighting, resulting in a dead-reckoned track that matched the locations determined from sightings with interpolated positions between them (Miller *et al.*, 2012). The correction vectors tended to be small, and in a consistent direction—indicating that the deviation between the dead-reckoned and sighting tracks arose due to water currents or offsets in the estimated speed of the whale not captured by using an average speed.

Horizontal speed and movement direction of the tagged whale were calculated from changes in the x-y location obtained from the corrected dead-reckoned track. Speed was calculated as the total great circle distance traveled over 2-min intervals. A dead-reckoned track could not be produced for record oo08_149a due to faulty readings in one axis of the accelerometer, so speed was calculated using the surfacing locations prior to and after each surfacing point. Direction of movement was calculated as the true bearing from the previous calculated position of the whale. Further details of the experimental procedure can be found in Miller *et al.* (2011, 2012).

B. Identification of avoidance responses

A specific set of objective criteria was used to determine whether or not an avoidance response was judged to have taken place. The primary criterion was to identify whether movement during the exposure period would cause the subject to move away from the source vessel or the source vessel path, and could therefore be judged to be avoidance. Inspection of the data sought to identify instances in which movement changed to avoidance, including changes in absolute travel direction, increases in the linearity of travel, or increases in

speed. We also inspected the data records to identify whether other behavioral patterns changed at the same time as the movement parameters, and considered other relevant factors that might have affected the movement of the whales such as bathymetry and social factors. For each record, we carefully inspected all data streams (Miller *et al.*, 2012), using the whale track to identify and describe broad scale changes in movement. The key movement parameters of interest were plotted for each exposure session (e.g., Fig. 2) and include speed, direction of movement relative to North, and direction of movement relative to the heading of the ship. Avoidance was judged to have occurred when movement during the exposure period was away from the source or perpendicular to the heading of the ship, as these would tend to move the animal away from the current position or future path of the ship. Increases in speed during avoidance movement trajectories were used as additional indicator of an avoidance reaction.

To aid interpretation of the movement data, we developed a quantitative indicator of behavioral change in order to evaluate whether changes in movement parameters observed during exposure sessions were unusual compared to baseline behavior periods before the first sonar exposure. We combined multivariate behavioral metrics using Mahalanobis distance to create a univariate metric that quantified differences in behavior between adjacent time periods. Direction of movement and speed data from corrected dead-reckoning tracks were sampled at 1 min intervals (interpolated to one minute intervals from track data for record oo08_149a). Direction of movement was decomposed into Northing as cosine of direction and Easting as sine of direction. The covariance matrix of speed, Northing, and Easting was calculated from the entire track record. A 10 min window was slid at 1 min steps across the entire data record, and we quantified the mean Mahalanobis distance (Manly, 2005) between the two sets of adjacent five 1 min data points for speed, Northing, and Easting. The maximum break-point value within the exposure session was noted and then compared to the maximum value within mock exposure periods randomly placed within the baseline period. Under a null hypothesis of no response, we would expect the maximum break-point values within the exposure session to fall within the range of those in the baseline period. We calculated the proportion of time periods during baseline of equivalent-duration that had maximum break-point values that equal or exceed the value observed during the exposure (Table I; Fig. 2).

All of the objective criteria including the break-point analysis were considered together to form the final judgment of whether an avoidance reaction took place during each exposure session period. The judgment was initially made by two independent groups that then reached a consensus (Miller *et al.*, 2012). If avoidance was judged to have occurred, the precise time point of its onset was identified using the sighting track in combination with the diving, orientation, and acoustic records of the Dtag (Fig. 2).

C. Measurement of the acoustic dose

Sonar signals recorded by the Dtag and a calibrated array of hydrophones towed by the observation vessel were extracted for detailed analyses of the sonar received level.

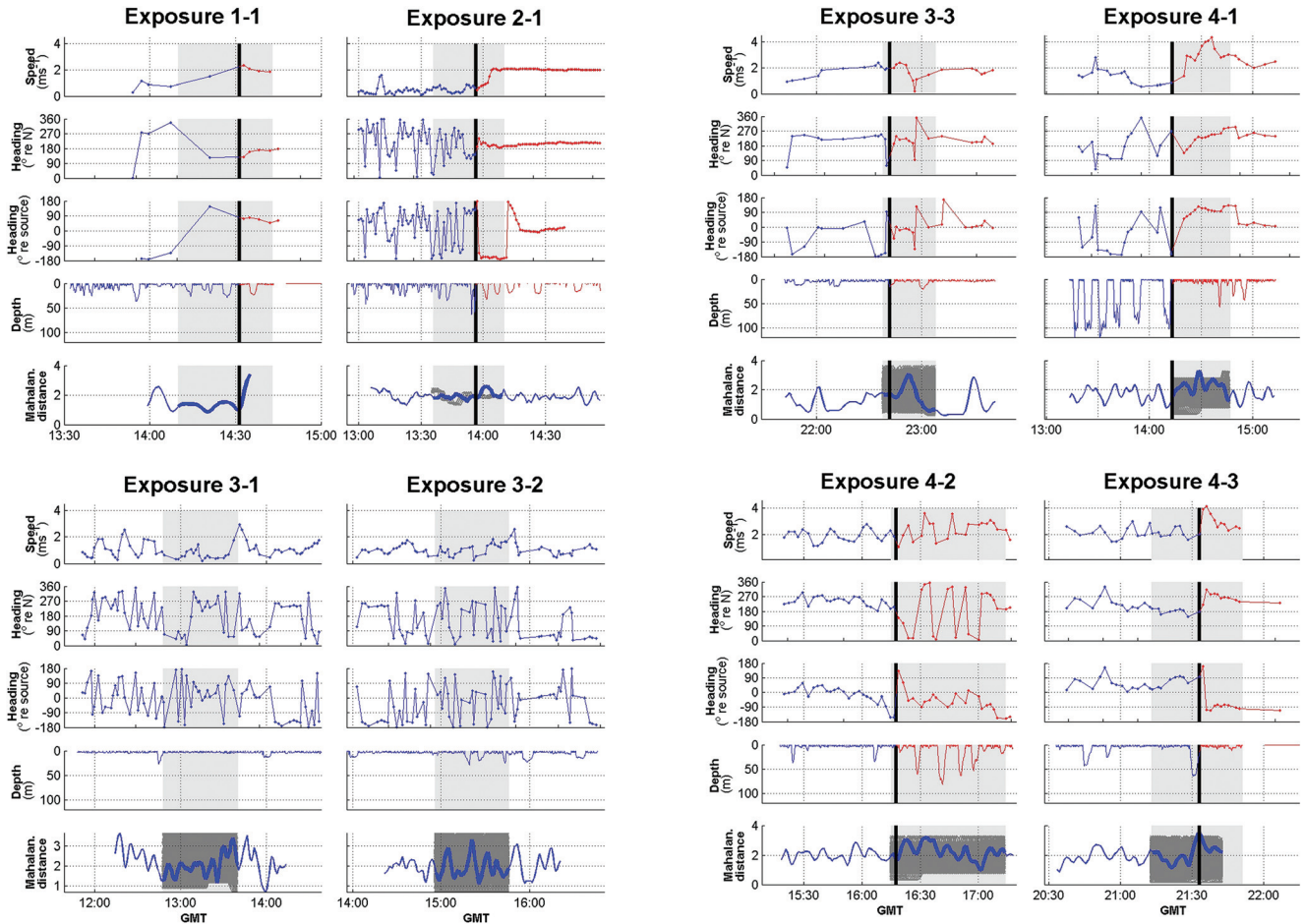


FIG. 2. Time series movement and diving data of tagged whales for all exposure sessions. All plots show a pre-exposure period, exposure (highlighted in gray, including the ramp-up period), and post-exposure periods for: speed over ground (top), heading of movement (second from top), and heading relative to course of the source vessel (third from top), dive depth (second from bottom), and Mahalanobis distance between the 5 min before and after each time step (bottom). In the bottom panel, Mahalanobis distance for the actual time series is shown as a blue line, while Mahalanobis distance for random intervals during baseline are shown as light gray lines. The vertical black lines indicate the time at which movement behavior was judged to have changed to avoidance. No line is shown for exposure sessions 3-1 and 3-2 as no avoidance was judged to have occurred in those cases.

The received levels for each ping were quantified as: (1) SPL_{max} , defined as the maximum SPL over a 200 ms averaging window for each sonar transmission; (2) sensation level, defined as the difference between the HT and the SPL of the received signal, and (3) SEL_{cum} , defined as the unweighted broadband cumulative SEL. Levels were calculated only when they exceeded noise levels recorded by the tag by 10 dB or more. The levels of some transmissions during source level ramp-up at the start of each exposure session were below this threshold. The received level of those pings was estimated as the measured level of the closest ping in time adjusted for the difference in source level.

The SPL of a sound wave of root-mean-square (RMS) acoustic pressure p_{RMS} was expressed in decibels (dB) relative to the standard reference pressure p_{ref} , which in water is one micropascal ($1 \mu Pa$; Morfey, 2001). In this paper the RMS averaging time was 200 ms, based on the hearing integration time of cetaceans and other mammals (Plomp and Bouman, 1959; Johnson, 1968).

SEL is a measure of received acoustic energy flux density. The SEL is the cumulative sum of squared pressures, and is mathematically described as

$$SEL = 10 \log_{10} \frac{\sum_{n=1}^N \int_0^T p_n^2 dt}{p_{ref}^2 t_{ref}}, \quad (1)$$

with $t_{ref} = 1$ s. This quantity is expressed in decibels with a reference value of $1 \mu Pa^2 s$. The SEL can be calculated for single noise events ($N = 1$) and multiple noise events ($N > 1$), both with individual events of T duration. Multiple-event SEL is commonly referred to as SEL_{cum} . Here SEL_{cum} was calculated from all received pings in the exposure session.

Sensation level was calculated as the difference between the received SPL and the HT. The HT was obtained by a non-linear fit to the mean behavioral audiogram data for three killer whales (Hall and Johnson, 1972; Szymanski et al., 1999), except ≥ 30 kHz where data from the Hall and Johnson (1972) study were ignored:

$$HT(f) = K - 20 \log_{10} \left(\frac{b^x f^x}{(a^x + f^x)(b^x + f^x)} \right) \quad (2)$$

where f is the frequency in hertz. The best-fit parameter values for K , a , b , and x were 27.26 dB re $1 \mu Pa$, 19 280 Hz,

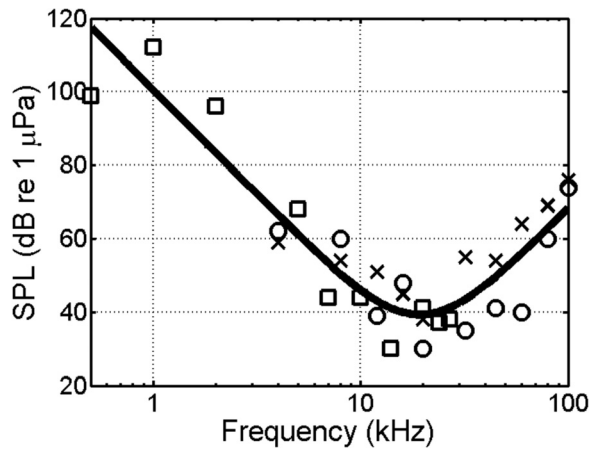


FIG. 3. Mean HT data available to date for three killer whales (squares: Hall and Johnson, 1972; circles and crosses: Szymanski *et al.*, 1999) and the killer whale audiogram derived for this paper. The weighting function used to calculate sensation level is the inverse of the audiogram values.

19 284 Hz, and 2.85, respectively (Fig. 3). This function is a modified version of the C- and M-weighting functions, but fitted with x as a free parameter instead of a value of 2 to match the steeper slopes of the hearing threshold (Wensveen *et al.*, 2014). The fitted HT (HT in dB re 1 μPa) was then inverted to obtain a weighting function (W in dB re 1 μPa^{-1}) as: $W(f) = -\text{HT}(f)$. The weighting function was applied in 1/3-octave bands with center frequencies between 1 and 40 kHz to calculate sensation level (Ellison *et al.*, 2012).

In exposure 2-1, tags were deployed on two different animals in the same group. These whales remained so close together that received levels for this experiment were determined from the average of the measurements from tags oo06_327s and oo06_327t. Similarly, in experiment 4 (exposures 4-1 to 4-3) tagged whale oo09_144b traveled consistently alongside oo09_144a. Acoustic measurements were not used from oo09_144b as measurements from the tag did not always correspond well with the measurements from oo09_144a or from the calibrated towed array. Based on extensive calibration of Dtags and recordings on the calibrated array, we are confident that our SPL measurements are accurate within ± 5 dB.

D. Calculating source-to-whale range using one-way travel time

The sound source closely follows the path of the source vessel at regular tow speeds and turning angles. The source's track was therefore similar to the ship's track but with a time delay caused by the length of the deployed tow cable (typically 100–200 m). The position of the source when pings were transmitted was derived from the track of the ship after correction for this time delay and the depth of the source.

The one-way travel time, or “time of flight” of the pings (time difference between transmission and arrival) was used to determine the source-to-whale range assuming an underwater sound speed of 1500 m/s. Ping transmission times were stored in UTC with high precision, but imprecise ping arrival times derived from the tag attachment time often created an offset in the range estimates. Using ordinary-least-squares, this offset was minimized by fitting the time-of-flight range

function to the range data derived from the whale sightings. The average ($N = 23$) RMS error of the fits was 80 m (range: 39–145 m), thus we consider ± 100 m to be a representative estimate of the uncertainty for the range measurements.

E. Specification of response thresholds for each exposure session

We used the observed response times to calculate the maximum dose received by the whale prior to the onset of avoidance, which was then considered to have been the response threshold for that exposure session. Response thresholds were derived and plotted for four different dose terms: SPL_{max} , sensation level, SEL_{cum} , and the proximity of the source to the whale. The acoustic thresholds for the observed responses were fitted to the dose-response function, detailed below. The proximity of the source at the time of the behavioral response was not fitted to a dose-response function because the range of parameter values for distance are very different than those for the acoustic dose term, but proximity as a dose term is reported and considered in the Discussion section.

In cases for which no response was judged to have occurred, we assume that the threshold was not reached for that animal during that session. These results are nonetheless informative; they tell us that the threshold for that exposure session was higher than the maximum dose received during the session. This is called “right censoring” (Plein and Moeschberger, 2003).

F. Estimating the cumulative dose-response function

We fitted the observed acoustic thresholds to a set of hierarchical Bayesian dose-response models. The “full” (i.e., most complex) model assumes that for any sonar exposure session, each whale (or whale group) has a response threshold that is a function of its typical average response threshold as well as two measured factors (previous exposure and sonar frequency band), and other sources of random between-session variation. We tested the importance of the two factors using a Bayesian model selection method (Gibbs Variable Selection, see below), and if they were not supported we also fitted more simple models that excluded one or both factors. All models are hierarchical in two senses. First they allow for variation in average threshold between whales, and also between individual exposure sessions within the same whale. Both of these are modeled as random effects. Second they separate the “process model,” which describes statistically the factors driving the true threshold of exposure for each exposure session, from the “observation model,” which links the true thresholds to the observed values, measured with error and in some cases right censoring. The hierarchy is shown in Fig. 4, and the elements further detailed below.

We favored a Bayesian formulation for the model because it allows flexibility in specifying the model, and also allowed us to specify priors on model parameters based on *a priori* knowledge of reasonable bounds for these parameters.

1. Process model

Let t_{ij} be the true, but unknown threshold of exposure that elicits a behavioral response for the i th whale on the j th

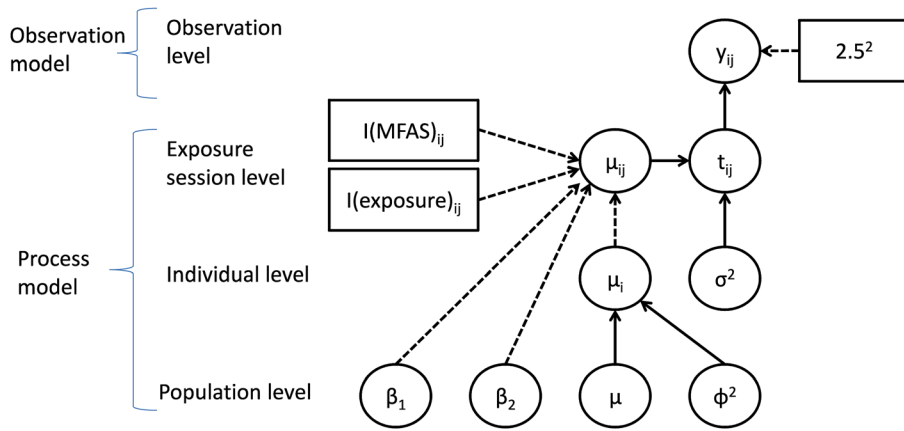


FIG. 4. (Color online) Directed acyclic graph showing the structure of the “full” hierarchical Bayesian model used to analyze the dose-escalation data. Variables in the model are represented by circles and constants by boxes. The variables are defined in Sec. II F of the text. Solid and dashed arrows indicate stochastic and deterministic relationships, respectively. Not shown are the constants required to define priors on the model parameters μ , ϕ , σ , β_1 , and β_2 —these are given in Table II. Reduced versions of the model omit β_1 , β_2 , or both.

exposure session. We assumed that this threshold follows a truncated normal distribution:

$$t_{ij} \sim TN(\mu_{ij}, \sigma^2, L, U), \quad (3)$$

where μ_{ij} is the expected threshold for the i th whale in the j th exposure session, σ^2 is the within-animal between-session variance in threshold, and L and U are a lower and upper limit to the threshold. Using truncation allowed us to incorporate assumptions about what constitutes a feasible range for the thresholds, and so constrained the dose-response function to lie within biologically reasonable bounds. For SPL_{\max} we assumed a lower limit of $L = 60$ dB re $1 \mu\text{Pa}$ and an upper limit of $U = 200$ dB re $1 \mu\text{Pa}$, for sensation level we assumed $L = 0$ dB and $U = 120$ dB, and for SEL_{cum} we assumed $L = 60$ dB re $1 \mu\text{Pa}^2 \text{ s}$ and $U = 200$ dB re $1 \mu\text{Pa}^2 \text{ s}$. The lower limit for SPL and SEL was based upon a conservative lower limit of detectability given hearing sensitivity and the lowest sea noise conditions; 0 dB for sensation level indicates that responses are assumed not to occur below the HT. The upper limit of 200 dB for SEL and SPL, and 120 dB for sensation level implies an assumption that all animals would begin to respond at those levels.

In the “full” model, we assumed that the expected threshold for the i th whale in the j th exposure session, μ_{ij} , depends upon the expected threshold for that whale, μ_i , as well as whether the whale has been exposed in the experiment before and the frequency band of the sonar signal used in the exposure session:

$$\mu_{ij} = \mu_i + \beta_1 I(MFAS)_{ij} + \beta_2 I(\text{exposed})_{ij}, \quad (4)$$

where β_1 is a parameter governing the effect of MFAS (Mid-frequency active sonar) relative to LFAS (Low-frequency active sonar), $I(MFAS)_{ij}$ is an indicator function that takes the value 1 if the exposure session was with MFAS (6–7 kHz band), 0 if LFAS (1–2 kHz band). The parameter β_2 represents the effect of previous exposure on threshold, and $I(\text{exposed})_{ij}$ is an indicator function that takes the value 0 for the first exposure session, 1 if the whale has been exposed in a previous sonar exposure session. Reduced versions of the model omitted the terms in Eq. (4) associated with β_1 and/or β_2 . Last, we assumed that the expected threshold for each whale follows a truncated normal distribution:

$$\mu_i \sim TN(\mu, \phi^2, L, U) \quad (5)$$

where μ is the mean threshold for all whales, ϕ^2 is the between-whale variance in threshold, and L and U are as defined above.

2. Input data and observation model

Each exposure session was scored as being “MFAS” when sonar in the 6–7 kHz band was used, and “LFAS” when 1–2 kHz sonar was transmitted. Sessions were also scored as being either the “first” exposure session, or an “exposed” session for the second and third exposure sessions conducted with animal groups 3 and 4 (Table I). In the cases where avoidance reactions were deemed to have occurred, response thresholds were quantified. To allow for uncertainty in the measured threshold, we assumed the measurement follows a normal distribution with a mean value of the true threshold, t_{ij} , and a standard deviation (s.d.) of 2.5 dB:

$$y_{ij} \sim N(t_{ij}, 2.5^2), \quad (6)$$

where y_{ij} is the measured threshold. Thus, the prior 95% density interval for the threshold was ± 5.0 dB around the point estimate, reflecting our confidence in the accuracy of the acoustic measurements.

In the cases where avoidance reactions were deemed not to have occurred, we only have a lower bound on the threshold—this being the maximum dose received. The true threshold must be higher than this value, and hence the lower bound on t_{ij} in these cases was set equal to the maximum SPL, sensation level or SEL received by the animal (depending on the dose term being modeled).

3. Specification of priors for the Bayesian models

Our goal was that the priors for the Bayesian model should be generally uninformative, while at the same time constraining the parameter estimates within biologically plausible bounds (Table II). Wide prior values were specified for μ , the mean threshold of an average whale that has not been exposed before and for an experiment involving 1–2 kHz sonar, which could take any value between 60 dB and 200 dB for SPL and SEL and between 0 and 120 dB for

TABLE II. Prior values used in the Bayesian model. Lower and upper limits are reported for uniform distributions (U), and mean and s.d. are reported for normal distributions (N).

Variable	SPL	Sensation level	SEL
μ	U(60,200)	U(0,120)	U(60,200)
σ	U(0,30)	U(0,30)	U(0,30)
ϕ	U(0,30)	U(0,30)	U(0,30)
β_1	N(0,30)	N(0,30)	N(0,30)
β_2	N(0,30)	N(0,30)	N(0,30)

sensation level with uniform probability. Priors for between and within animal standard deviation in response threshold (ϕ and σ , respectively) were set to be from a uniform distribution from 0 to 30 dB. The upper value of 30 dB for both parameters was chosen as it implies a 95% credibility interval of 120 dB, a similar range as that over which all thresholds were assumed to fall (60–200 dB for SPL and SEL or 0–120 dB for sensation level). Prior values for the influence of sonar frequency band (β_1) and previous exposure (β_2) were set as normal distributions with mean of zero and a large standard deviation (30 dB).

G. Application of the model to simulated data

We undertook a small simulation study to assess the ability of the model (and priors) to estimate a “true” dose-response relationship from a data set with a small sample size. The simulated data precisely replicated the number of trials, sonar frequency and order of exposure that we obtained in our real experiments. The mean threshold per whale (μ_i) in the simulation was sampled from a logistic function similar to the dose-response function used by the U.S. Navy for its environmental assessments (U.S. Navy, 2008). The logistic function used was $P = 1/[1 + \exp(165.8 - RL)/4.6]$, where P is the probability of response, and RL is the received SPL. The thresholds at which each whale responded in each simulated exposure session depended upon its mean threshold, the influence of sonar frequency and prior exposure, and within-animal variation. These were sampled from a truncated normal distribution with a mean given by Eq. (6) and s.d. (equivalent to σ) of 5 dB. In summary, the parameter values used in the simulation were $\mu = 165.8$ dB, $\phi = 8$ dB, $\sigma = 5$ dB, $\beta_1 = 10$ dB, $\beta_2 = -5$ dB. Importantly, these values differed from the mean values of the priors used in estimation. Also, the shape of the dose response function differed from that assumed by the priors of the model, providing a further test of the robustness of the method. We fitted 1000 simulated datasets drawn from this parameter set, and compared the distribution of posterior estimates of the parameters with the actual simulated values.

H. Procedure for fitting and interpreting the Bayesian model

Model fitting was performed using a Markov chain Monte Carlo (MCMC) algorithm, implemented using the software JAGS 3.2.0 via the rjags package (Plummer, 2011) in R 2.14.1 for Mac OS X (R Development Core Team, 2011). Burn-in (i.e., number of samples required from

initialization to convergence so that further samples come from the posterior distribution) was found to be rapid. Results are based on 100 000 samples, with each parameter started at its prior mean, after a burn-in of 10 000.

To assess the level of support to retain the β terms in the final dose-response model, we applied Gibbs Variable Selection (GVS) as detailed by Ohara and Sillanpaa (2009), on the full model. This method calculates the probability (GVS p -value) that a given variable’s inclusion in the model is supported by the data or not, with zero indicating no support for inclusion and 1.0 indicating full support for inclusion. In cases where there was not strong support for inclusion of a beta term, we dropped the corresponding term from the model and re-fitted the model. We (arbitrarily) defined “not strong support” to mean a GVS p -value < 0.95 , but the results are not sensitive to any reasonable level used as in practice all fitted values were found to be < 0.6 .

III. RESULTS

We conducted a total of eight sonar exposure sessions with four different killer whale groups in which at least one whale was tagged (Table I). Two whales were tagged simultaneously for two of these experiments. For these two cases, tracking was systematically conducted on one focal animal (oo06_327s and oo09_144a) with only opportunistic sightings of the second tagged whale. Only the data from the focal animal were analyzed here, because the whales were always seen in close proximity, and we do not assume the behavior of different individuals within a group to be independent. Experiments oo06_317s and oo06_327s were conducted during November while whales were seen feeding upon overwintering herring. Due to short daylight hours in 2006, only one sonar exposure was conducted with these groups. Experiment oo08_149a was conducted in late May within a narrow fjord that connects to Vestfjord. No feeding was observed. A total of three exposure sessions were conducted with this group (6–7 kHz, 1–2 kHz, and 6–7 kHz). Experiment oo09_144a was conducted offshore Vesterålen and the whales were producing tail-slaps and other feeding related sounds during dives recorded before the start of the sonar exposure (Simon *et al.*, 2005). Again three exposure sessions were conducted with this group (1–2 kHz, 6–7 kHz, and 1–2 kHz downsweep).

A. Description of behavior and determination of response for each exposure session

Here we describe the outcome of each exposure session, referring to data plotted in Fig. 2. An example geometry plot is shown for exposure session 2–1 (Fig. 1). Plots showing geometry, time-series data, and acoustic propagation models for all exposure sessions are available in Miller *et al.* (2011).

1. Experiment 1, oo06_317s

Exposure 1-1. This whale was within a large group of 50–80 animals that were carousel feeding upon herring. The dive profile and sound recordings on the tag indicated

that the tagged animal changed from feeding to travel behavior shortly before the start of the ramp-up transmissions that started the exposure session. The source vessel at that point was more than 6 km from the whale (Miller *et al.*, 2011), so the change was not likely due to the presence of the vessel itself. The playback vessel approached from behind the direction of motion of the group, which weakens our ability to detect any turn away from the approaching vessel. The whales continued to move northeast, until the subject and its group made a gradual turn toward southeast, which coincided with an increase in swimming speed (Fig. 2). This change in movement direction is consistent with avoidance of the source vessel. The Mahalanobis distance statistic was inconclusive, likely because the turn judged to be avoidance was quite gradual, and changes in movement parameters calculated using the Mahalanobis distance indicated that changes were generally greater in the pre-exposure period, when the tagged animal was mostly carousel feeding, than during the exposure period. Whales circle herring when carousel feeding and produce strong fluking movements when tail-slapping prey, leading to strong changes in movement parameters during this behavioral context. The precise time of the increase in speed was identified from the flow noise recorded on the tag, which changes with swimming speed. Flow noise increased by more than 6 dB at 14:31:30 UTC, which was used as the point at which avoidance was judged to start. The speed continued to be high until the tag detached prematurely from the animal, which caused us to stop the exposure. No post-exposure data are available as without the tag we were unable to continue following the animal.

2. Experiment 2, oo06_327s and oo06_327t

Exposure 2-1. Two individuals were tagged within a carousel feeding group (Fig. 1), with feeding indicated by recording of tail-slap sounds (Simon *et al.*, 2005) in the early part of the record (see Miller *et al.*, 2011), which ceased abruptly during the exposure period. Before the change point the whales were moving at low speed with a highly tortuous travel path (Figs. 1 and 2). There was a clear change in behavior at the end of a synchronous deep dive (14:56:14-14:56:25). All indications of feeding ceased, and the animals' movement path became highly directional, with an increase in speed. Despite the erratic movement patterns in the pre-exposure baseline period, the Mahalanobis break-point statistic indicated that the maximum change during the exposure period was greater than in any similar duration period in the pre-exposure baseline period. The animals continued to move Southwest in the direction out of Vestfjord. The tags were recovered 28 and 30 km away 4–5 h later, indicating an extended period of high speed travel.

3. Experiment 3, oo08_149a

Exposures 3-1 and 3-2. This killer whale group contained a calf and was found within a relatively narrow fjord. No photo-id or acoustic matches were made with herring-feeding whales from the winter within Vestfjord. This group, which may have attacked a minke whale prior to our arrival,

was followed for 1.5 days before a tag was successfully attached to a small female or adolescent male. Before the tag was attached, the whales milled for a long period of time in the lower part of the narrow fjord, and made one return trip to the head of the fjord. The calf was regularly observed during this period, and was never observed alone. In the pre-exposure baseline period, the whales again milled in shallow water in the bottom end of the fjord. During the first two exposures, the animals continued to mill in shallow water in the bottom end of the fjord (Fig. 2, Miller *et al.*, 2011). The Mahalanobis distance statistic was inconclusive, indicating that maximum change intervals in the exposure periods were no greater than that in the baseline period. Thus, both exposures 3-1 and 3-2 were scored as no response.

Exposure 3-3. Before this exposure, the whales moved to the narrow head of the fjord and turned back toward the southwest, as they had done once during the pre-tagging period. The whales made a clear change of direction during a long dive in ramp-up, which resulted in their crossing to the eastern side of the fjord. The movement responses during this exposure were likely constrained by the whales' location in this narrow part of the fjord. The whales also increased speed immediately after the same dive (Fig. 2), which increased their distance from the source ship from 0.5 to 1.2 km, based on time-of-flight analysis of the sonar signal (Miller *et al.*, 2011). Later during the exposure period, the group moved southeast at a slower speed, and the source ship came closer to the group. High frequency whistles (Samarra *et al.*, 2010) were produced by the tagged whale group during the 6–7 kHz sonar transmissions. The calf was seen traveling alone in the same direction of travel, more than 1000 m behind the group near the end of this relatively slower-moving period, but it is not known precisely when the separation first occurred. This was the first time that the calf was seen traveling alone over 2.5 days of observing the group. A mitigation stop to the sonar was called, but the transmissions had actually just stopped anyway following the timing of the experimental protocol. The observation vessel followed the calf as part of the mitigation protocol, which limited our ability to track the tagged whale. The calf was always oriented toward the rest of the group when it was observed at the surface. Following the end of the sonar transmissions, the calf later rejoined the group, after traveling alone for at least 86 min. In post-exposure, which lasted 7 h after the final sonar ping, this group was observed to return to the position they occupied during the first two exposures in the lower end of the fjord, and the calf was seen in close proximity to other group members.

Interpretation of the outcome of this exposure session is complicated by the narrow fjord and the calf separation, both of which could have a strong influence on the movement behavior of the group. The Mahalanobis distance analysis was inconclusive with maximum change values being consistently higher during the pre-exposure baseline period (when the animals were milling in the bottom of the fjord) than during the exposure period. Applying a precautionary judgment, we concluded that the original turn to cross the fjord, which coincided with an initial increase in speed, was the onset of avoidance during the session. The movement direction was

subsequently constrained by the narrowness of the fjord, and the group speed reduced due to the separation of the calf. We used the change in one of the axes of the magnetometer in the Dtag during the dive as a response change point.

4. Experiment 4, oo09_144a and oo09_144b

Exposure 4-1. This group of killer whales was located offshore, but both photo-id and acoustic call matches were made with killer whales sighted feeding on herring inshore during winter months. Two adult males were tagged in the group, and animal oo09_144a was chosen as the focal whale. All members of the group were closely associated throughout the follow. The tagged animals were making deep dives during the pre-exposure period, with tail-slaps recorded, indicating active feeding. The pre-exposure surface track indicates slow horizontal movement with regular changes in direction. Whale A had just surfaced from a deep dive prior to the first ping transmitted. Some whale calls were recorded after the first ping, most clearly on tag B, as tagged whale A surfaced at this time. More coordinated calling started after the second ping, and a highly coordinated and strong vocal response started after ping 3, and escalated into many loud calls consistently being produced immediately after each ping. The two tagged whales became more synchronous in their dive patterns and the group swam in a lined-up configuration with decreased group spacing (Miller *et al.*, 2011). No further indications of foraging were apparent after the start of the exposure session. The animals' movement was judged to be a very clear example of avoidance, with a strong and sustained increase in speed, increase in the directionality of movement, and movement consistently perpendicular to the path of the source vessel (which turned to continue to approach the tagged group; Fig. 2). The Mahalanobis distance statistic indicated that the maximum change value during exposure was greater than any during the pre-exposure baseline period. The acoustic response started very early in the exposure period. Tagged animals A and B went on a longer surface dive and animal A clearly increased fluking stroke magnitude between the second and third pings. Whale B seemed to have increased fluking motions immediately after the first ping, although this increase was less strong than that from whale A. We therefore marked the onset of avoidance as the start of increased fluking after the second ping at 14:13:30, which was also associated with the longer-duration surface dive.

Exposure 4-2. During the pre-exposure interval for 4-2, the animals were still traveling following their response to the first 1–2 kHz exposure, though at a normal speed. A change in direction consistent with avoidance occurred early in this 6–7 kHz exposure session (Fig. 2). The group turned away from the source, and also made sideways turns perpendicular to the path of the source vessel during which they increased their speed. Near the point of closest approach, the whales turned east, 90° away from the approach path of the source ship, and sped up. The whales then turned to continue to travel in the southwesterly direction in which they had been going before this exposure. The Mahalanobis distance statistic indicated an unusual change in movement, but was

inconclusive in this case, with 29% of randomly selected periods during baseline having maximum change values exceeding those observed during the actual exposure. The onset of avoidance was judged to occur at the time of the turn away from the source, at 16:17:40.

Exposure 4-3. Prior to this exposure, the whales continued to travel in a southwesterly direction following the first two exposures. An increase in speed and change in direction indicate an avoidance reaction after the deep dive in full exposure to the 1–2 kHz downsweep (Fig. 2). The whales then maintained a course perpendicular to the approach path of the source, later returning to their southwesterly course. The maximum change value in the Mahalanobis distance was greater than any in the pre-exposure baseline period. The time of the maximum change value (21:33:10) was taken as the time of the turn away from the source. Tag oo09_144a detached just after this final exposure, but tag oo09_144b remained attached for an additional 101 min during which time some surface indications of feeding were observed.

In summary, avoidance behavior was determined to have occurred in six of the eight sonar exposure sessions (Table I). Though the specific form varied from exposure to exposure, avoidance reactions were typified by increases in speed, a change in direction of movement or a change to a more linear direction of movement, and/or movement correlated to the path of the source vessel (Fig. 2). A consistent feature we observed as part of all avoidance responses was horizontal movement perpendicular to the course of the approaching source vessel.

B. Determining the response thresholds

For all sessions in which a response was judged to have occurred, the acoustic received levels and distance at the onset of the response were determined (Table I; Fig. 5). There was a high level of correlation between the different dose terms at the onset of the avoidance responses (Fig. 5).

C. Fitting the acoustic response thresholds to the dose-response model

We calculated separate dose-response functions for the three different acoustic response threshold dose terms. For SPL_{max} the posterior mean estimate for β_1 (effect of sonar frequency) in the full model was a substantial -19 ± 18 dB, indicating a trend for lower response thresholds during MFAS (6–7 kHz) exposure sessions than LFAS (1–2 kHz) sessions (Table III). However, Gibbs Variable Selection indicated little support for inclusion of sonar frequency in the final model (GVS $p = 0.51$). There was also little support for inclusion of a term encoding an effect of previous sonar exposure (β_2 ; GVS $p = 0.39$). Hence both terms were excluded from the final model. Final parameter values are shown in Table III, and the corresponding dose-response curve in Fig. 6. The results for SEL_{cum} (Table III) were very similar to those for SPL_{max} , which was expected for a dose-escalation design as the SEL_{cum} is strongly influenced by the escalating SPL_{max} values.

When sensation level was used as the dose term, there was no indication of any effect in the full model of either

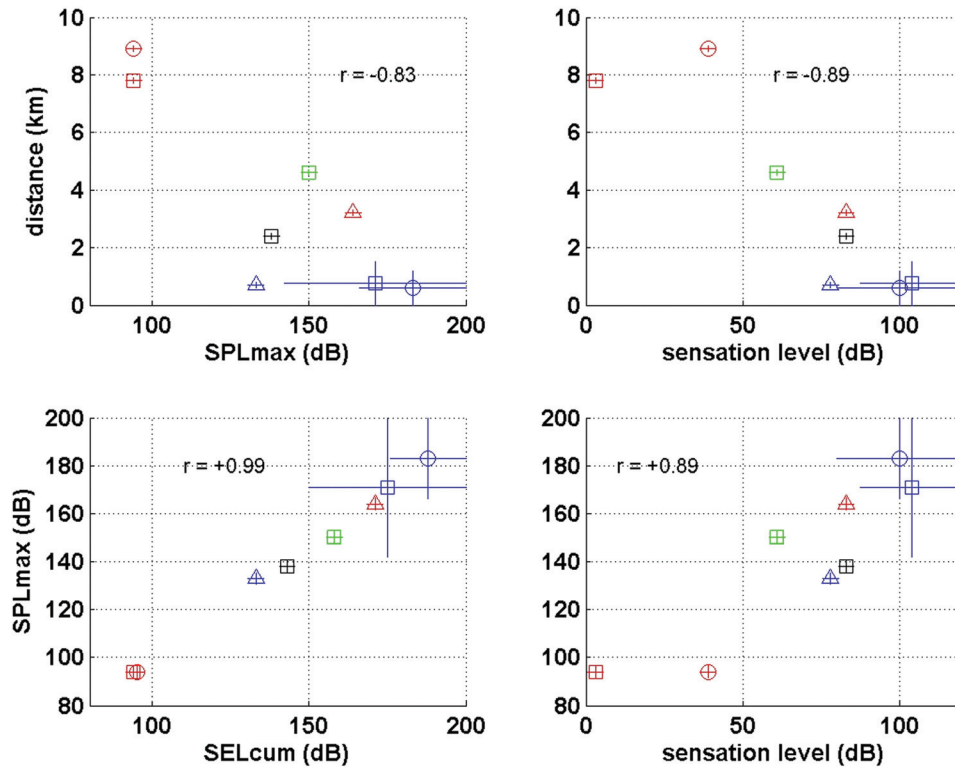


FIG. 5. Relationship of response thresholds for different ways of representing the sonar dose. The symbol is plotted on the mean values, and the horizontal and vertical blue lines indicate the lower and upper limits of the predicted response threshold for the cases where no avoidance behavior was detected. Symbol color indicates the experiment id (green: exp. 1, black: exp. 2, blue: exp. 3, red: exp. 4), while the symbol shape indicates the order in which the session fell within the experiment with square, circle and triangle being exposure sessions 1, 2, and 3, respectively. The correlation coefficient (r) is reported for each pair of values. Decibel reference values are $1 \mu\text{Pa}$ (SPL_{max}), 1 (sensation level), and $1 \mu\text{Pa}^2 \text{ s}$ (SEL_{cum}).

order of exposure or sonar frequency (Table III), and this finding was reflected in low Gibbs p -values. This indicates that the response thresholds for *LFAS* (1–2 kHz) and *MFAS* (6–7 kHz) sonar signals differed less once differences in hearing sensitivity at those frequencies were taken in to account. The dose-response curve excluding the β terms is shown in Fig. 7.

TABLE III. Posterior mean estimates \pm s.d. for parameters following model fitting to the data. The top 7 rows show model posteriors to the full model including estimates for the β terms. GVS refers to the Gibbs Variable Selection parameter calculated for each β term. The final three rows show the model posterior estimates after model selection, for which none of the β terms were retained. The final column shows a sensitivity analysis for SPL values excluding oo08_149a session 3, which was conducted in a narrow fjord in which avoidance may not have been as well detected.

Variable	SPL_{max} (dB re $1 \mu\text{Pa}$)	Sensation level (dB)	SEL_{cum} (dB re $1 \mu\text{Pa}^2 \text{ s}$)	SPL_{max} excluding exposure 3-3
Full model (β terms included)				
μ	150 ± 20	64 ± 22	157 ± 20	147 ± 21
σ	25 ± 4	26 ± 3	26 ± 3	25 ± 4
ϕ	18 ± 8	20 ± 8	19 ± 8	20 ± 7
β_1	-19 ± 18	5 ± 19	-21 ± 18	-16 ± 19
β_2	8 ± 18	14 ± 19	8 ± 18	15 ± 20
Variable selection				
β_1 GVS p	0.51	0.38	0.54	0.47
β_2 GVS p	0.39	0.46	0.40	0.45
After final model selection (no β terms included)				
μ	142 ± 15	74 ± 17	149 ± 16	147 ± 17
σ	26 ± 3	25 ± 3	26 ± 3	26 ± 3
ϕ	16 ± 8	19 ± 8	17 ± 8	18 ± 8

For all three acoustic representations of the dose, the posterior parameter estimates of the final model (which did not include any β terms) for both within- and between-whale variation were high (Table III), suggesting that unaccounted factors other than the sonar frequency and previous exposure drive a high level of variation in response thresholds. Nonetheless, in all cases, the posterior dose-response curve had smaller credible intervals than the prior, reflecting the relevance of the information contained in the data despite the small sample size.

D. Fitting simulated thresholds to the dose-response model

Our simulation test confirmed that the dose-response model, with the combination of specified priors and limited data observations available in a dataset matching our true dataset, can estimate an underlying dose-response relationship with a minimal degree of bias. The peak of the distribution of posterior estimates for the parameters closely matched the simulated values (Fig. 8). The simulated parameter $\mu = 165.8 \text{ dB re } 1 \mu\text{Pa}$ was estimated as $168.7 \text{ dB re } 1 \mu\text{Pa} \pm 5.7 \text{ dB}$ (mean \pm s.d.) across the 1000 simulations. The parameter $\phi = 8 \text{ dB}$ was estimated as $12.1 \pm 3.5 \text{ dB}$. The parameter $\sigma = 5 \text{ dB}$ was estimated as $10.1 \pm 3.9 \text{ dB}$. The parameter $\beta_1 = 10 \text{ dB}$ was estimated as $11.7 \pm 5.7 \text{ dB}$. The parameter $\beta_2 = -5 \text{ dB}$ was estimated as $-4.5 \pm 5.1 \text{ dB}$.

IV. DISCUSSION

Our study utilized multi-sensor observations of whale movements and behavior to identify avoidance reactions of killer whales during experimentally controlled exposures to sonar signals. We inspected data of each exposed whale,

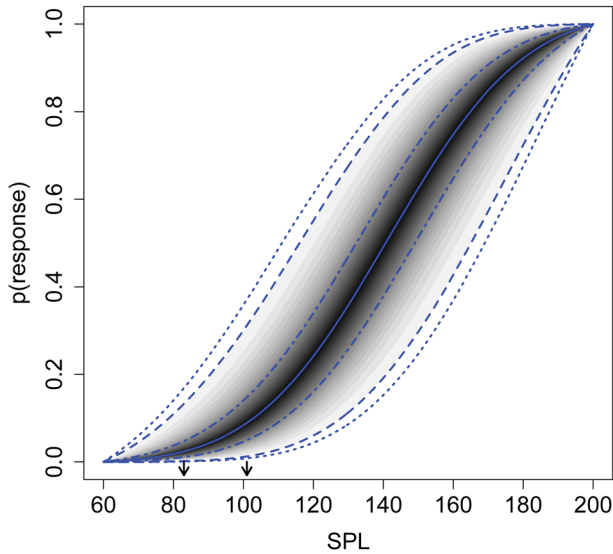


FIG. 6. (Color online) Posterior dose-response curve showing the probability of onset of avoidance against received SPL (dB re 1 μ Pa). The solid central line represents the mean, followed by 50%, 95%, and 99% credible interval lines (see also Table IV). Note that the dose-response model assumes the signal is audible over the range, but the limited data on the threshold of hearing for 1–2 kHz signals by killer whales [Eq. (2)] indicates that sensitivity ranges from 101 dB re 1 μ Pa at 1 kHz to 83 dB re 1 μ Pa at 2 kHz (marked in the figure with small arrows).

identified the time for onset of avoidance, and quantified the response thresholds using three different acoustic dose terms as well as proximity to the source (Table I). In our experiments, these different dose terms for the response thresholds tended to be correlated with each other (Fig. 5). The acoustic thresholds were then fitted to a Bayesian dose-response model which provides initial estimates of population level avoidance response thresholds, between-animal and within-animal variation in response thresholds, and the extent to which sonar frequency or previous exposure within the experimental sequence affected response thresholds. A

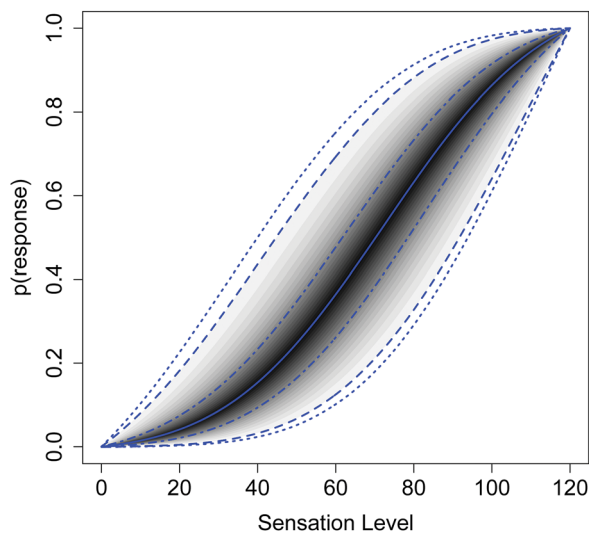


FIG. 7. (Color online) Posterior dose-response curve showing the probability of onset of avoidance against received sensation level (SPL minus HT). The solid central line represents the mean, followed by 50%, 95%, and 99% credible interval lines.

simulation study showed that the model was capable of estimating a true underlying dose-response function with little bias given the specified priors, and structure and size of our dataset. Here we discuss these results and consider how the dose-response functions should be of immediate use to managers wishing to assess the environmental risk sonar might pose to the behavior of cetaceans (Boyd *et al.*, 2008). Finally, we link our experimental results to published observations of killer whales in Vestfjord during an actual navy sonar exercise in 2006.

A. Predictions of the dose-response model and experimental results

The clearest conclusion to be drawn from the posterior parameter estimates is that, after testing for the possible influence of sonar frequency and previous exposure, a high level of unexplained between (ϕ) and within (σ) individual variability in avoidance response thresholds was still apparent. A similarly high level of within and between animal variation in thresholds was estimated for all of the acoustic dose terms (Table III). High levels of variation in response thresholds indicate that contextual variables or unexplained differences between individuals, such as previous experiences with anthropogenic sounds, had a profound impact on response thresholds in our study, consistent with conclusions of previous studies (Southall *et al.*, 2007; Ellison *et al.*, 2012).

An influence of sonar frequency on response thresholds was not supported in the Gibbs Variable Selection procedure, but overall, SPL thresholds were 19 ± 18 dB lower for 6–7 kHz than 1–2 kHz exposure sessions. A 19 dB difference in SPL_{\max} response thresholds is close to the 20–30 dB difference in auditory sensitivity measured for this species (Fig. 3). In contrast, there was no indication of any effect of sonar frequency when sensation level was used as the dose term (Table III). Weighting received SPL by the HT curve has been suggested to be appropriate for estimating susceptibility of disturbance by noise when loudness data are unavailable (Finneran and Schlundt, 2011). However, our data are not conclusive on this point as there is a high degree of uncertainty (18 dB s.d.) around the mean difference (19 dB). Our study does not provide adequate support to justify inclusion of a frequency effect in the dose-response function. However, such an effect should not be ruled out and could only be more conclusively evaluated with larger sample sizes.

There was little consistent influence of previous exposure in a sonar session for any of the acoustic dose terms (Table III). This result is concordant with the outcomes of the two experiments in which subjects were exposed to more than one sonar session (Table I). Subject oo08_149a was judged to start avoidance at a received SPL_{\max} of 133 dB re 1 μ Pa in its third trial, after not responding during the first two exposures when it was exposed to SPL_{\max} of 142 and 166 dB in the first and second exposure sessions, respectively. In contrast, subject oo09_144a started avoidance at a much higher SPL_{\max} threshold of 164 dB re 1 μ Pa in its third exposure, than during the first two exposures with SPL_{\max} avoidance thresholds of 94 dB re 1 μ Pa. The third exposure session

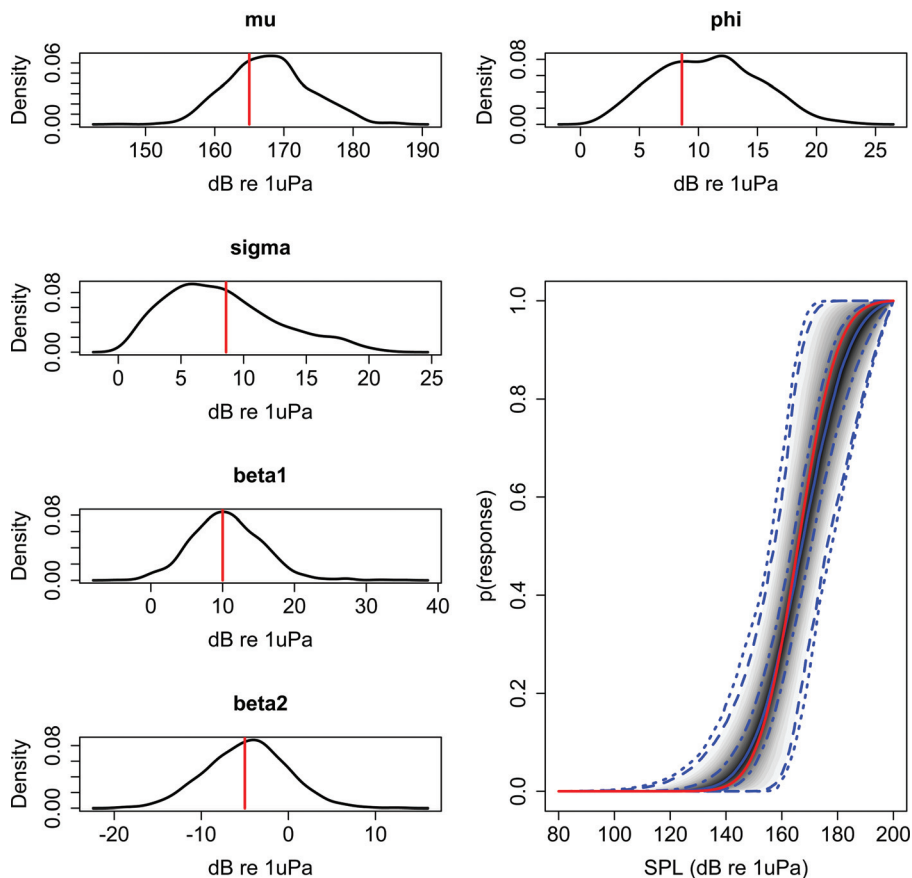


FIG. 8. Posterior densities of the Bayesian dose-response function parameters for a typical simulation using the U.S. Navy dose-response curve for odontocetes other than beaked whales and harbor porpoises (red line in all panels) as the underlying dose-response function, and a data sample equivalent to that obtained in our study. Note that the posterior distributions were centered upon, or only slightly different from, the simulated parameter values: $\mu = 165$ dB re 1 μ Pa, $\phi = 8$, $\sigma = 5$ dB, $\beta_1 = 10$ dB, $\beta_2 = -5$ dB. The bottom-right panel illustrates the posterior dose-response function (black) with 50%, 95%, and 99% credible intervals and the simulated dose-response function derived from the U.S. Navy curve (red).

for oo09_144a was a downsweep, instead of an upsweep (Table I), which also may have influenced the response threshold (Miller *et al.*, 2012; Kastelein *et al.*, 2013).

Given the lack of a consistent influence of either sonar frequency or whether or not a subject had previously been exposed, the resulting diversity of response thresholds are estimated as arising from unexplained between (ϕ) and within (σ) individual variability in avoidance response thresholds. The dose-response functions were strongly influenced by the very low avoidance thresholds documented in experiment oo09_144a (Table I). The response threshold of 94 dB re 1 μ Pa during the 1–2 kHz exposure corresponded to a sensation level threshold of just 3 dB. As noted in Table I, this response was associated with a large number of calls being produced immediately after each ping. The acoustic response gives us additional confidence that the signal was audible to at least some of the individuals in the group, even at such low sensation levels.

Our experiments indicate that avoidance reactions to sonar may be associated with negative biological consequences. Higher swimming speed during avoidance entails increased energetic cost of locomotion, and the transition from feeding to avoidance travel may result in decreased feeding opportunities. Before sonar exposures 2-1 and 4-1 we recorded social calling and tail-slap sounds, which indicate feeding on herring (Simon *et al.*, 2005). No clear indications of feeding were heard after the avoidance reaction started during these exposures, but some social calling was

observed 4.75 and 0.5 h after the final respective exposures 2-1 and 4-3. The most severe response we observed in association with avoidance responses was the temporary separation of a calf from its group in exposure 3-3 which was the only occasion it was seen alone over a 62.2 h observation period. This separation triggered a mitigation stop to our experimental exposure, and the calf rejoined its group 65 min after sonar transmissions stopped. Social disruption could be more severe in longer duration sonar exercises with multiple sources that also might use higher source level sonars than were used in our experimental exposures (see Sec. IV B).

B. Application of the results to operational sonar use

An important goal of our study was to derive dose-response relationships that could be used to predict behavioral responses of cetaceans to sonar sounds. However, even though our experimental exposures were conducted using a real naval sonar source care should be taken in extrapolating the results of our experiments to predict outcomes of real naval sonar trials that take place over longer time periods, possibly also with more ships involved. Source ships in real exercises likely move in more random directions relative to the positions of whales than the source ship did in our short-duration experimental exposures, during which the animal was intentionally approached by the source vessel. Our experiments should be quite representative for animals that happen to be ahead of moving sonar ships, but response

TABLE IV. Mean, median, 50% and 95% credible interval limits for the final SPL_{max} dose-response curve (Fig. 6).

SPL_{max} received level dB re 1 μ Pa	Mean	Median	25%	75%	2.50%	97.50%
60	0.00	0.00	0.00	0.00	0.00	0.00
70	0.01	0.01	0.00	0.01	0.00	0.05
80	0.03	0.02	0.01	0.04	0.00	0.11
90	0.06	0.04	0.02	0.08	0.00	0.20
100	0.10	0.08	0.05	0.14	0.01	0.30
110	0.17	0.15	0.09	0.23	0.03	0.43
120	0.26	0.25	0.17	0.34	0.06	0.56
130	0.37	0.36	0.26	0.47	0.12	0.68
140	0.49	0.49	0.38	0.60	0.20	0.79
150	0.62	0.63	0.52	0.73	0.30	0.87
160	0.74	0.76	0.66	0.84	0.44	0.94
170	0.83	0.85	0.78	0.91	0.58	0.97
180	0.91	0.92	0.88	0.96	0.73	0.99
190	0.96	0.97	0.95	0.99	0.87	1.00
200	1.00	1.00	1.00	1.00	1.00	1.00

thresholds to ships moving away from a whale could differ from what we observed in our study.

The dose-response functions assume that responses can occur over the range set by the priors (60–200 dB re 1 μ Pa for SPL), but it is only realistic to predict behavioral responses to occur at levels that are at least audible to the whales. HT at 6–7 kHz of roughly 55 dB re 1 μ Pa are compatible with the SPL dose-response function presented here (Fig. 6; Table IV), but the available audiogram data indicate that HT for the 1–2 kHz band are in the range of 83–101 dB re 1 μ Pa. The curve presented for sensation level (Fig. 7) references SPL to the HT and is therefore an efficient way to deal with the influence of audibility on the probability of response. However, audibility could also potentially be limited by masking caused by ambient noise sources. Our results indicate that some responses appear to start at received levels low enough to be close to the limits of audibility, so the effects of audibility and masking are important to consider in future studies. It is particularly relevant to consider whether signal to noise ratio itself is an important predictor of response as is indicated in some studies (Dunlop *et al.*, 2013). We recommend that application of the dose-response functions derived here to evaluation of potential effects of real exercises should take into account the question of audibility of the signals in question, and set response probabilities to zero when they clearly would not be audible.

We attempted to make our experiments representative of real sonar exercises, using an actual naval towed source capable of producing a high source level (197–214 dB re 1 μ Pa m), but some operational sonars operate at even higher source levels. If received SPL is the dose term that truly predicts animal response, differences in source level lead in general (depending on sound propagation conditions) to corresponding differences in the amount of habitat over which animals would be expected to respond. In our experiments, responses started about 0.7–8.9 km from the source vessel (Table I). The mean observed proximity at the start of avoidance was 3.8–4.6 km for all exposure sessions, with a mean

SPL threshold of 142 dB re 1 μ Pa (Table III). To extrapolate the distances that might be associated with those received levels for a sonar with source level equal to 225 dB re 1 μ Pa m in the two frequency bands tested, we calculated propagation loss versus distance using spherical spreading plus frequency-dependent absorption (Urlick, 1975, p. 102) and separately using cylindrical spreading (Ainslie, 2010, p. 467, Eq. 9.80) with $D=1$ assuming a surface duct of thickness $H=100$ m and vertical gradient of sound velocity $c' \equiv dc/dz = 0.016$ /s. For the 6–7 kHz band, a received SPL of 142 dB is predicted to occur at distances of 9 km for spherical spreading and 24 km for cylindrical spreading, while for the 1–2 kHz band a SPL of 142 dB is predicted to occur at 12 and 245 km, respectively. For a 225 dB re 1 μ Pa m sonar source level, the estimated distances to the received level equivalent to the 50% SPL response threshold could therefore be roughly three to as much as 60 times further than the distances (proximity) that actually occurred during our experiments. Though many studies report the received SPL associated with behavioral changes in marine mammals (Southall *et al.*, 2007; Miller *et al.*, 2012), it is unknown whether acoustic received level or proximity to the sonar (Table I) are the most effective predictors of avoidance response, or are the most relevant features of the signal to which animals respond. DeRuiter *et al.* (2013) quantified behavioral indices of Cuvier's beaked whales (*Ziphius cavirostris*) during separate exposures from distant operational sonar exposures and nearby experimental exposures, and found that both distance and received level were important predictors of response intensity. We cannot distinguish between these possibilities on the basis of our experiments alone. This continued gap in knowledge is important because differences in source levels from those used in experimental studies such as ours and exercises using higher source levels alter the relationship between the received level and proximity to the source vessel. However, consistent with the finding of DeRuiter *et al.* (2013), we might predict that for the same received sound level, whales will be less likely to avoid more distant sources than nearby ones as closer proximity of any hazard is likely to be perceived as an additional risk factor. Most likely, behavioral response thresholds will vary as a combined function of both received level and proximity, and the context in which an exposure occurs will likely modify the influence of different features of the sonar dose (Ellison *et al.*, 2012). More observations during actual exercises are needed to determine the amount of habitat over which cetaceans behaviorally respond to operational sonar use and the duration and extent of avoidance reactions over more realistic time scales than in our short experimental exposures which were designed to illuminate thresholds at which avoidance reactions start.

Kuningas *et al.* (2013) analyzed how sonar activity might have impacted killer whale presence in the Vestfjorden basin. Whale numbers sighted in Vestfjord gradually declined after the start of the 2006 naval fleet FLOTEX Silver exercise in Vestfjorden. The avoidance behaviors we observed and the SPL dose-response relationships for onset of avoidance of sonar we derived from our experimental exposures with killer whales in this area are consistent with the conclusion of

Kuningas *et al.* (2013) that killer whales left the area to avoid sonar activity. In our experiments, we observed animals moving away from sonar sources over ranges of several km, and whales often moved many km away from the location of sound exposure. Source levels of the sonars used in the FLOTEX trial were likely to have been higher than those used in our experiments, but we have no means to estimate what levels were received by the whales. High levels of between and within animal variability in response thresholds noted as an important outcome of our experiments (Table III) are consistent with the gradual decline in whale numbers over the first days of the exercise as less sensitive individuals possibly remained in the area longer than more sensitive individuals.

C. Methods considerations, simulation and sensitivity analyses

We applied a data-analysis approach used in Phase-I clinical trials in medicine with humans. Phase-I clinical trial results are considered to be preliminary information, often using a small number of patients and dose-escalation to derive initial estimates of dose functions (Simon *et al.*, 1997; O'Quigley and Conaway, 2010). The approach uses expert judgment to identify whether a given response occurred and to then fit the observed thresholds to an assumed, underlying dose-response model. The benefit of adopting this approach for our study is that data from free-ranging animals can be used to derive initial estimates of dose-response functions despite the limited sample-sizes typically achieved in at-sea experiments. The importance of the simulation result was to demonstrate that the Bayesian modeling approach with associated priors was able to recover an underlying dose-response function with limited bias based upon the size of the dataset we were able to obtain in the real experiments. This gives us confidence that the posterior estimates of the dose-response model are not overly influenced by the priors, but rather reflect trends in the data. Inspection of posterior distributions indicate that specific estimates of both within and between whale variation were somewhat constrained by the prior distribution of these parameters (a uniform distribution with support from 0 to 30 dB). Thus, it is possible that these parameters may have been underestimated and that the true between- and within-animal variability was even somewhat greater than what was derived in our posterior estimates. However, more data would be needed to better quantify values for between and within-animal variability in response thresholds, and to attempt to identify specific factors which drive this high level of variability.

Following the approach in phase-I clinical trials, we examined each experimental exposure session in detail to determine whether an avoidance reaction occurred and the time of the onset of avoidance. Though experts used objective measures from the data in their judgments about whether or not avoidance occurred, the final decision was based upon expert judgment and not on a specific quantitative criterion. The small sample sizes relative to the variation in behavior patterns, and lack of matching experimental control data,

limited our ability to test statistically against a null hypothesis or to specify to which feature of the sonar exposure animals actually responded. Given the distances from the vessel at which avoidance reactions started (Table I), we do not think it is likely that propulsion noise from the ship alone was the driver of the avoidance reactions. Killer whales in the study area have been repeatedly observed to be approached closely by fishing and whale-watch vessels without obvious reactions such as those we observed in our experiments. Furthermore, the maneuvering of the observation vessel relative to the whale was kept constant across all experimental conditions, but its proximity might have changed the responsiveness of the whales to sonar. Herring schools present in the same area as these experiments did not respond to sonars at higher received levels (Doksæter *et al.*, 2009), so it is unlikely that the whales were simply following responses of herring.

To aid in the expert evaluation of the outcome of each experiment, we applied a break-point analysis to each experimental exposure session to calculate a metric of how likely changes in movement observed during exposure sessions could have been caused by chance given movement patterns before the sonar exposure. The break-point analysis indicated an unusual change during exposure periods for three exposure sessions, all which were judged to have been avoidance responses. Of the five sessions for which the break-point statistic was inconclusive, two records (oo08_149a sessions 1 and 2) showed no indications of avoidance in the track or in the time-series data plots and were considered not have contained any avoidance behavior. Two of the other three records (oo06_317s and oo09_144a session 2) had clear indications of movement away from the source vessel or the source vessel path along with increases in speed. In these cases, the break-point was inconclusive due to limited pre-exposure data (oo06_317s) or due to an extended set of movement changes (oo09_144a session 2) rather than a single clear change in movement.

The most problematic case was oo08_149a session 3, which was unusual in several respects. The break-point statistic was inconclusive, but the track and time-series data plots show indications of brief movement away from the source when the whales crossed to the other side of the narrow fjord. It is conceivable that the whales would have continued to move sideways away from the source path had they not been constrained by bathymetry. An initial increase in speed near the start of the exposure session was followed by a period of decreased speed, during which the calf was sighted traveling alone. Later the focal group increased speed again while the calf was still separated from its group. We concluded that a precautionary interpretation of the data was that the whales began to respond by moving away from the source at higher speed at a received SPL of 133 dB re 1 μ Pa, but decreased speed to allow the calf to rejoin the group. However, as the indications of avoidance specifically are the weakest in this case, we conducted a sensitivity analysis, and refit the dose-response model leaving out this exposure session as the narrow fjord context reduced the ability to detect avoidance. The results (Table III, right column), indicate that removing this case does not substantially

change the parameter estimates or the general conclusions about the influence of sonar frequency or order effect. Fortunately, the avoidance reactions that most greatly influence the dose-response function (experiment 4; oo09_144a) are also the clearest examples of avoidance in our dataset.

V. CONCLUSIONS

Though our sample size of experiments was small, the dose-response curves derived from our data provide experimentally supported indications of the risk of behavioral response of free-ranging killer whales to sonar. The curve we derived differs substantially from that used by the U.S. Navy in its environmental impact assessments for this species, particularly at received SPL below 165 dB re 1 μ Pa (U.S. Navy, 2008, 2012). The differences between our results and those behind the U.S. Navy curve could result from differences in responsiveness of the different species tested, differences in methods, or from differences in context, such as captive vs wild settings or feeding activity or not. In general, the more similar the experimental setup is to the actual operational situation, the more confident one can be about extrapolating from experimental data to naval operations. This would suggest weighting data from wild animals exposed to realistic moving sonar sources more heavily than data from captive animals exposed to simulated sonar sounds from a stationary source nearby. Interestingly, a behavioral response dose-response function for captive bottlenose dolphins before habituation was more similar to the dose-response function derived here than after the dolphins apparently habituated to the sound exposure (Houser *et al.*, 2013). Southall *et al.* (2007) argued that some species (e.g., porpoises and beaked whales) may be particularly sensitive to anthropogenic sound, as reflected by a recent U.S. Navy environmental impact statement (U.S. Navy, 2012). It is possible that our free-ranging killer whales were more sensitive than the captive bottlenose dolphins and captive beluga whales in the Navy data (Finneran and Schlundt, 2004; Houser *et al.*, 2013). Our research group has collected similar experimental data on long-finned pilot (*Globicephala melas*) and sperm (*Physeter macrocephalus*) whales, which may be less sensitive than the killer whales reported here. Clear responses at received SPL <100 dB re 1 μ Pa were not observed in those other species (Miller *et al.*, 2012), as they were for killer whale experiment oo09_144a in this study.

The success of our approach in determining a dose-response function using a realistic sonar exposure for wild cetaceans, even with a small sample size, suggests that such research can be an important basis for managing risk of anthropogenic sound to marine mammals. Though care is needed in extrapolating from our experimental results to predict effects of actual sonar exercises, we found the predictions from our dose-response functions were consistent with an observed decrease in whale numbers in a whale watching area during a FLOTEX sonar exercise in 2006 (Kuningas *et al.*, 2013). Avoidance reactions of killer whales in our experiments were associated with cessation of feeding and a calf separation. While our short experiments did not harm

the subjects directly, consequences such as cessation of foraging or the separation of a calf from its group, as revealed in our experiments, could pose a risk if sonar exposure was prolonged, more intense or cumulated with other noise sources. Our analysis also revealed that a high degree of unexplained within and between animal variation existed in response thresholds. Such variation suggests that other factors (e.g., density of prey, exposure history, behavioral context) not controlled for in our study might strongly influence whale movements and responsiveness to sonar (Goldbogen *et al.*, 2013; Ellison *et al.*, 2012). Such variation is likely to affect responsiveness of cetaceans to real sonar exercises, as well. Therefore, in addition to a minimum number of controlled dose-response experiments, we strongly recommend observations of animal responses during actual sonar exercises, along with ongoing ecosystem-based monitoring to understand more completely the sources of fluctuations in whale numbers in different locations.

ACKNOWLEDGMENTS

We thank the science and ships' crews of the RV *HU Sverdrup II* and MS *Strønstad* for the effort during data collection at sea. This study was funded by the U.S. Office of Naval Research to the 3S and MOCHA projects, the Royal Norwegian Navy and the Norwegian Ministry of Defense, the Netherlands Ministry of Defense, and by WWF Norway. Visual data were collected using Logger 2000, developed by the International Fund for Animal Welfare (IFAW) to promote benign and non-invasive research. The authors acknowledge the support of the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) in the completion of this study. MASTS is funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions. Thanks to R. Dekeling and J. Finneran for helpful review of the manuscript. Animal experiments were carried out under permits issued by the Norwegian Animal Research Authority (Permit No. 2004/20607 and S-2007/61201), in compliance with ethical use of animals in experimentation. The research protocol was approved by the University of St Andrews Animal Welfare and Ethics Committee and the WHOI Institutional Animal Care and Use Committee.

- Ainslie, M. A. (2010). *Principles of Sonar Performance Modeling* (Springer-Praxis, Chichester, UK), pp. 800.
- Boyd, I. L., Brownell, R., Cato, D., Clarke, C., Costa, D., Evans, P., Gedamke, J., Gentry, R., Gisiner, Gordon, J., Jepson, P., Miller, P., Rendell, L., Tasker, M., Tyack, P., Vos, E., Whitehead, H., Wartzog, D., and Zimmer, W. (2008). "The effects of anthropogenic sound on marine mammals: A draft research strategy." European Science Foundation—Marine Board, position paper 13.
- Buck, J. R., and P. L. Tyack. (2000). "Response of gray whales to low frequency sounds," *J. Acoust. Soc. Am.* **107**, 2774.
- Cisneros-Montemayor, A. M., Sumaila, U. R., Kaschner, K., and Pauly, D. (2010). "The global potential for whale watching," *Mar. Policy* **34**, 1273–1278.
- Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., D'Amico, A., D'Spain, G., Fernández, A., Finneran, J., Gentry, R., Gerth, W., Gulland, F., Hildebrand, J., Houser, D., Hullar, T., Jepson, P. D., Ketten, D., MacLeod, C. D., Miller, P., Moore, S., Mountain, D., Palka, D., Ponganis, P., Rommel, S., Rowles, T., Taylor, B., Tyack, P., Wartzok, D., Gisiner, R.,

- Mead, J., and Benner, L. (2006). "Understanding the impacts of anthropogenic sound on beaked whales," *J. Cetacean Res. Manage.* **7**, 177–187.
- D'Amico, A. D., Gisiner, R. C., Ketten, D. R., Hammock, J. A., Johnson, C., Tyack, P. L., and Mead, J. (2009). "Beaked whale strandings and naval exercises," *Aquat. Mamm.* **35**, 452–472.
- DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., Thomas, L., and Tyack, P. L. (2013). "First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar," *Biol. Lett.* **9**: 20130223.
- Doksæter, L., Godø, O. R., Handegard, N. O., Kvadsheim, P. H., Lam, F. P., Donovan, C., and Miller, P. J. O. (2009). "Behavioral responses of herring (*Clupea harengus*) to 1–2 and 6–7 kHz sonar signals and killer whale feeding sounds," *J. Acoust. Soc. Am.* **125**, 554–564.
- Dunlop, R. A., Noad, M. J., Cato, D. H., Kneist, E., Miller, P. J. O., Smith, J. N., and Stokes, M. D. (2013). "Multivariate analysis of behavioural response experiments in humpback whales (*Megaptera novaeangliae*)," *J. Exp. Biol.* **216**, 759–770.
- Ellison, W. T., Southall, B. L., Clark, C. W., and Frankel, A. S. (2012). "A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds," *Conserv. Biol.* **26**, 21–28.
- Finneran, J. J., and Schlundt, C. (2004). "Effects of intense pure tones on the behavior of trained odontocetes," Technical Report 1913, Spawar Systems Center, San Diego.
- Finneran, J. J., and Schlundt, C. E. (2011). "Subjective loudness level measurements and equal loudness contours in a bottlenose dolphin (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **130**, 3124–3136.
- Goldbogen, J. A., Southall, B. L., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Hazen, E. L., Falcone, E. A., Schorr, G. S., Douglas, A., Moretti, D. J., Kyburg, C., McKenna, M. F., and Tyack, P. L. (2013). "Blue whales respond to simulated mid-frequency military sonar," *Proc. R. Soc. B* **280**, 20130657.
- Hall, J. D., and Johnson, C. S. (1972). "Auditory thresholds of a killer whale *Orcinus orca* Linnaeus," *J. Acoust. Soc. Am.* **51**, 515–517.
- Houser, D. S., Martin, S. W., and Finneran, J. J. (2013). "Exposure amplitude and repetition affect bottlenose dolphin behavioral responses to simulated mid-frequency sonar signals," *J. Exp. Mar. Biol. Ecol.* **443**, 123–133.
- Johnson, C. S. (1968). "Relation between absolute threshold and duration-of-tone pulses in the bottlenosed dolphin," *J. Acoust. Soc. Am.* **43**, 757–763.
- Johnson, M. P., and Tyack, P. L. (2003). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," *IEEE J. Oceanic Eng.* **28**, 3–12.
- Kastelein, R. A., Gransier, R., van den Hoogen, M., and Hoek, L. (2013). "Brief behavioral response threshold levels of a harbor porpoise (*Phocoena phocoena*) to five helicopter dipping sonar signals (1.33 to 1.43 kHz)," *Aquat. Mamm.* **39**, 162–173.
- Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (1982). *Fundamentals of Acoustics* (Wiley, New York), pp. 496.
- Kuningas, S., Kvadsheim, P. H., Lam, F.-P. A., and Miller, P. J. O. (2013). "Killer whale presence in relation to naval sonar activity and prey abundance in northern Norway," *ICES J. Mar. Sci.* **70**, 1287–1293.
- Lusseau, D., Bain, D. E., Williams, R., and Smith, J. C. (2009). "Vessel traffic disrupts the foraging behaviour of southern resident killer whales *Orcinus orca*," *Endang. Species Res.* **6**, 211–221.
- Madsen, P. T., Johnson, M., Miller, P. J. O., De Soto, N. A., and Tyack, P. L. (2006). "Quantitative measures of air-gun pulses recorded on sperm whales (*Physeter macrocephalus*) using acoustic tags during controlled exposure experiments," *J. Acoust. Soc. Am.* **120**, 2366–2379.
- Manly, B. F. J. (2005). *Multivariate Statistical Methods: A Primer*, 3rd ed. (Chapman and Hall/CRC, Boca Raton, FL), pp. 214.
- McCawley, R. D., Fewtrell, J., Duncan, A. J., Jenner, C., Jenner, M.-N., Penrose, J. D., Prince, R. I. T., Adhitya, A., Murdoch, J., and McCabe, K. (2000). "Marine seismic surveys—A study of environmental implications," *APPEA J.* **40**, 692–708.
- Miller, P. J. O., Antunes, R., Alves, A. C., Wensveen, P., Kvadsheim, P., Kleivane, L., Nordlund, N., Lam, F.-P. A., van IJsselmuide, S., Visser, F., and Tyack, P. L. T. (2011). The 3S experiments: studying the behavioural effects of naval sonar on killer whales (*Orcinus orca*), sperm whales (*Physeter macrocephalus*), and long-finned pilot whales (*Globicephala melas*) in Norwegian waters: Scottish Oceans Institute Technical Report SOI-2011-001. <http://soi.st-andrews.ac.uk/documents/424.pdf> (Last viewed 10/09/12).
- Miller, P. J. O., Johnson, M. P., Madsen, P. T., Biassoni, N., Quero, M., and Tyack, P. L. (2009). "Using at-sea experiments to study the effects of air-guns on the foraging behavior of sperm whales in the Gulf of Mexico," *Deep-Sea Res. Pt. I* **56**, 1168–1181.
- Miller, P. J. O., Johnson, M. P., Tyack, P. L., and Terray, E. A., (2004). "Swimming gaits, passive drag and buoyancy of diving sperm whales *Physeter macrocephalus*," *J. Exp. Biol.* **207**, 1953–1967.
- Miller, P. J. O., Kvadsheim, P., Lam, F.-P. A., Wensveen, P. J., Antunes, R., Alves, A. C., Visser, F., Kleivane, L., Tyack, P. L., and Sivle, L. D. (2012). "The severity of behavioral changes observed during experimental exposures of killer (*Orcinus orca*), long-finned pilot (*Globicephala melas*), and sperm whales (*Physeter macrocephalus*) to naval sonar," *Aquat. Mamm.* **38**, 362–401.
- Morfe, C. L. (2001). *Dictionary of Acoustics* (Academic, San Diego, CA), pp. 430.
- Morton, A. B., and Symonds, H. K. (2002). "Displacement of *Orcinus orca* (L.) by high amplitude sound in British Columbia, Canada," *ICES J. Mar. Sci.* **59**, 71–80.
- NMFS (2005). "Assessment of acoustic exposure on marine mammals in conjunction with USS SHOUP active sonar transmissions in the eastern Strait of Juan de Fuca and Haro Strait, 5 May, 2003" (NOAA Office of Protected Resources, Washington, D.C.), report dated January 21, 2005.
- Nowacek, D. P., Johnson, M. P., and Tyack, P. L. (2004). "North Atlantic right whales (*Eubalaena glacialis*) ignore ships but respond to alerting stimuli," *Proc. R. Soc. B* **271**, 227–231.
- Nowacek, D. P., Thorne, L. H., Johnston, D. W., and Tyack, P. L. (2007). "Responses of cetaceans to anthropogenic noise," *Mammal Rev.* **37**, 81–115.
- O'Hara, R. B., and Sillanpää, M. J. (2009). "A review of Bayesian variable selection methods: What, how and which," *Bayesian Anal.* **4**, 85–118.
- O'Quigley, J., and Conaway, M. (2010). "Continual reassessment and related dose-finding designs," *Stat. Sci.* **25**, 202–216.
- O'Quigley, J., Pepe, M., and Fisher, L. (1990). "Continual reassessment method: a practical design for phase-I clinical trials in cancer," *Biometrics* **46**, 33–48.
- Plein, J. P., and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd ed. (Springer, New York), pp. 536.
- Plomp, R., and Bouman, M. A. (1959). "Relation between hearing threshold and duration for tone pulses," *J. Acoust. Soc. Am.* **31**, 749–758.
- Plummer, M. (2011). "rjags: Bayesian graphical models using MCMC. R package version 3-5," <http://CRAN.R-project.org/package=rjags> (Last viewed 10/09/12).
- R Development Core Team (2011). "R: A language and environment for statistical computing. R Foundation for Statistical Computing," Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/> (Last viewed 10/09/12).
- Richardson, W. J., Greene, C. R. J., Malme, C. I., and Thomson, D. H. (1995). *Marine Mammals and Noise* (Academic, San Diego, CA), pp. 576.
- Samarra, F. I. P., Deecke, V. B., Vinding, K., Rasmussen, M., Swift, R. J., and Miller, P. (2010). "Killer whales (*Orcinus orca*) produce ultrasonic whistles," *J. Acoust. Soc. Am.* **128**(5), EL205–EL210.
- Simon, M., Whalberg, M., Ugarte, F., and Miller, L. A. (2005). "Acoustic characteristics of underwater tail slaps used by Norwegian and Icelandic killer whales (*Orcinus orca*) to debilitate herring (*Clupea harengus*)," *J. Exp. Biol.* **208**, 2459–2466.
- Simon, R., Freidlin, B., Rubenstein, L., Arbuck, S. G., Collins, J., and Christian, M. C. (1997). "Accelerated titration designs for phase I clinical trials in oncology," *J. Natl. Cancer I* **89**, 1138–1147.
- Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., Richardson, W. J., Thomas, J. A., and Tyack, P. L. (2007). "Marine mammal noise exposure criteria," *Aquat. Mamm.* **33**, 411–521.
- Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (1999). "Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioural audiograms," *J. Acoust. Soc. Am.* **106**, 1134–1141.
- Tyack, P. L., Zimmer, W. M. X., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. W., Clark, C. W., D'Amico, A., DiMarzio, N., Jarvis, S., McCarthy, E., Morrissey, R., Ward, J., and Boyd, I. L. (2011). "Beaked whales respond to simulated and actual navy sonar," *PLoS One* **6**(3), e17009.

- Urick, R. J. (1975). *Principles of Underwater Sound for Engineers*, 2nd ed. (McGraw-Hill, New York), pp. 342.
- U.S. Navy. (2008). "Southern California range complex, final environmental impact statement/overseas environmental impact statement," December 8, 2008. <http://www.socalrangecomplexeis.com/Documents.aspx> (Last viewed 10/30/13).
- U.S. Navy. (2012). "Draft environmental impact statement/overseas environmental impact statement for Atlantic fleet training and testing," May, 2012. <http://afiteis.com/DocumentsandReferences/AFTTDocuments/DraftEISOEISMay2012.aspx> (Last viewed 10/09/12).
- Wensveen, P. J., Huijser, L. A. E., Hoek, L., and Kastelein, R. A. (2014). "Equal latency contours and auditory weighting functions for the harbour porpoise (*Phocoena phocoena*)," *J. Exp. Biol.* (in press).
- WWF-Norway (2001). "Lack of environmental assessment resulted in negative impact on fish and tourism," Letter from WWF-Norway to Norwegian Minister of Defense, MOD archive no 01/01309-1 (in Norwegian).