J Exp Biol Advance Online Articles. First posted online on 15 November 2012 as doi:10.1242/jeb.071498
Access the most recent version at http://jeb.biologists.org/lookup/doi/10.1242/jeb.071498

Multivariate analysis of behavioural response experiments in humpback whales (*Megaptera novaeangliae*)

Rebecca A. Dunlop[1], Michael J. Noad[1], Douglas H. Cato[2,3], Eric Kniest[4], Patrick Miller[5], Joshua N. Smith[6] and M. Dale Stokes[7].

[1] *School of Veterinary Science, University of Queensland, Gatton, Qld 4343, Australia*

[2] *Defence Science and Technology Organisation, Eveleigh, NSW 1430, Australia*

[3] *University of Sydney Institute of Marine Science, University of Sydney, NSW 2006, Australia*

[4]*School of Engineering, University of Newcastle, NSW 2308, Australia*

[5]*Scottish Oceans Institute, East Sands, St Andrews, Fife, KY16 8LB. United Kingdom.*

[6] *School of Biological Sciences and Biotechnology, Murdoch University, WA 6150, Australia*

[7]*Marine Physical Laboratory, Scripps Institution of Oceanography, Calif 92037-0238, USA.*

Running title: A humpback whale BRS

Correspondence: [1] *School of Veterinary Science, University of Queensland, Gatton, Qld 4343, Australia*. r.dunlop@uq.edu.au

Key words: behavioural response; BRS; social sounds; anthropogenic noise; humpback whales

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Copyright (C) 2012. Published by The Company of Biologists Ltd

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

SUMMARY

The behavioural response study (BRS) is an experimental design used by field biologists to determine the function and /or behavioural effects of conspecific, heterospecific or anthropogenic stimuli. When carrying out these studies in marine mammals it is difficult to make basic observations and achieve sufficient samples sizes due to the high cost and logistical difficulties. Rarely are other factors such as social context or the physical environment considered in the analysis because of these difficulties. This paper presents results of a BRS carried out in humpback whales to test the response of groups to one recording of conspecific social sounds and an artificially-generated tone stimulus. Experiments were carried out in September/October 2004 and 2008 during the humpback whale southward migration along the east coast of Australia. In total, 13 ‘tone’ experiments, 15 ‘social sound’ experiments (using one recording of social sounds) and three silent controls were carried out over two years. Results (using a mixed model statistical analysis) suggested that humpback whales responded differently to both stimuli, measured by changes in course travelled and dive behaviour. Although the response to ‘tones’ was consistent, in that groups moved offshore and surfaced more often (suggesting an aversion to the stimulus), the response to ‘social sounds’ was highly variable and dependent upon the composition of the social group. The change in course and dive behaviour in response to ‘tones’ was found to be related to proximity to the source, the received signal level and signal-to-noise ratio *SNR*. This study demonstrates the complexity of behavioural responses by a marine mammal to acoustic stimuli, the need to replicate stimuli to generate a sufficient sample size, and to measure as many other factors as possible culminating in a multivariate analysis in order to tease out complex interactions.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

INTRODUCTION

Acoustic playback experiments, ‘behavioural response studies’ (BRS) and ‘controlled exposure experiments’ (CEE) seek to identify and describe potential responses of animals to natural or synthetic acoustic stimuli. Playback experiments (in which animal sounds are played back) have been carried out since the late 1950s on a variety of species including insects (Alexander, 1961; Haskell, 1957), birds (Ficken and Ficken, 1970; Roche, 1966; Verner and Milligan, 1971), fish (Fish, 1968), seals (Watkins and Schevill, 1968), reindeer (Espmark, 1971) and cetaceans (Morgan, 1970). However, there is a distinct lack of published playback experiments in marine mammals compared to terrestrial animals. In the last five years alone, over 200 playback experiments have been carried out on bird species (to 2006) compared to only 46 experiments on marine mammals (Deecke, 2006). Since that report (to date), only a few more playback studies on a marine mammal species have been presented in the peer-reviewed literature. This difference in the number is likely due to the relative ease with which wild terrestrial animals can be targeted or experimental terrestrial animals can be held in captivity (Falls, 1992) compared to most marine mammal species. There are obvious logistical and monetary constraints in marine mammal research (which limits the sample size and therefore the experimental power, Dunlop et al., 2012) and there is a lack of background data on marine mammal populations available to test hypotheses and interpret conclusions. More recently, the terms CEE and BRS have been used for experiments that control the acoustic dosage (level received by the animal), with exposure metrics measured or modeled at the animal, usually in order to obtain dose-response. Many playback experiments do not include this level of control.

Humpback whales are very vocal. Males produce a long, complex, stereotyped, repetitive ‘song’ (Payne and McVay, 1971) on the breeding grounds and during migration. The function of song is currently under debate but one function is likely to be a as a sexual advertisement directed towards females (Smith et al., 2008; Tyack, 1981). Other proposed functions include song operating as a mechanism for male social sorting (Darling et al. 2006), a method of spacing between singers (Frankel et al., 1995) or a threat display during intra-sexual competition (Baker & Herman, 1984). In addition to song, humpback whales produce ‘social sounds’ (Payne, 1978; Tyack, 1981), which include surface-generated percussive sounds (e.g. breaches, pectoral flipper slaps, tail slaps) and social vocalisations. Social sounds are produced by adult males, adult

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

females (Dunlop et al., 2008) and probably calves (Zoidis et al., 2008). These sounds are thought to convey information on species and sex of the signaller, signaller location, size, readiness to mate and readiness to compete with males as well as aid in group cohesion during joining, instigate and facilitate social interactions between groups or cohorts, maintain contact with other group members and facilitate group splitting (Dunlop et al., 2008). However, the function of specific social sounds is still not known and the contextual use of many of the social sounds in humpback whales is still to be determined.

To date, two ‘playback’ studies have been carried out on humpback whales, both designed to determine the function of conspecific vocalisations; Tyack (1983), tested the behavioural response of humpback whales to conspecific song and social sounds, and Mobley Jr. et al. (1988), included exposure to a synthetic sound along with playback of conspecific song and social sounds. Both singers and non-singers demonstrated approach and avoidance responses to playback of social sounds (Tyack, 1983; Mobley et al., 1988) suggesting an important communicative function of these sounds between different social groups. However, as with many of these studies, sample size was unavoidably small and the experiments were ‘sacrificially replicated’ (Deecke, 2006), that is, focal individuals were used repeatedly (exposed to both stimuli) and statistical independence was violated in the analysis as it did not account for this repeated measure design. Other behavioural response studies on humpback whales have focused on assessing the response to an anthropogenic stimulus. In the marine mammal literature they are usually referred to as controlled exposure experiments (CEEs). In many of these studies, only the received level of the sound was considered as the stimulus variable and other factors relating to the context of the exposed animal (for example the social environment and the noise environment) were not considered. The most current and preferred term for this type of experiment in wild cetaceans is ‘Behavioural Response Study’, or ‘BRS’.

In this experiment, we used a typical behavioural response experimental design to test the response of humpback whales to one recording of conspecific social sounds compared to a low-frequency sweep (2 kHz) tone, which is within the frequency range of humpback vocalisations. Song units are highly variable in frequency range and usually lie between 30 Hz (Payne and Payne, 1985) and 4 kHz (Tyack and Clark, 2000) with harmonics extending beyond 24 kHz (Au et al., 2006) and social vocalisations range from less than 30Hz to 2.5 kHz (Dunlop et al., 2007).

We therefore assumed both stimuli were audible but hypothesised that humpback whale groups would react differently to an artificial signal ('tones') compared to a more natural signal (a recording of conspecific social sounds taken from the same population of whales). Behavioural responses to a sound stimulus are likely to be context-specific, both in terms of the social context of the animal and the context of the source stimulus tested (signal-to-noise level, proximity of the source, novelty of the source). Therefore, we used a multivariate analysis to test for an effect of categorical factors such as social context and the presence of other cohorts such as singing whales (the social environment) and continuous variables such as received signal-to-noise level, proximity to the source vessel and background noise levels (the external environment) on the behavioural response to each stimulus type.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

MATERIALS AND METHODS

Study site

Initial experiments were carried out in September/October 2004 during the humpback whale south-ward migration. Further experiments were carried out in 2008 during the same two months. The study site was located at Peregian Beach, which is 150 km north of Brisbane, on the east coast of Australia (26°29'S, 153°06'E) and about 800km south of the potential breeding grounds in the Great Barrier Reef (Smith et al., 2012). Humpback whales passing Peregian Beach are migrating from the breeding grounds further north and show a range of behaviours typical of breeding grounds (for example singing, forming competitive groups, frequent joining and splitting of groups, meandering and variation in swim speed and direction, nursing and other maternal behaviours due to numerous new born calves) while moving in a general southwards direction. A fixed array of hydrophones was moored offshore for acoustic data collection. Each hydrophone was suspended from a buoy which transmitted the acoustic data to a base station on shore. Buoys 1-3 were 1.5 km from the beach, parallel to the shoreline and approximately 0.7 km apart. Buoys 4 and 5 extended seaward from buoy 2 in a line perpendicular to the shore and were approximately 0.5km apart. Buoys 1 - 3 were always operational and were usually adequate to fix the positions of vocalising whales (using Ishmael software, Mellinger 2001). This was supplemented with buoys 4 and 5 for many observations. Visual survey teams were based on an elevated survey point, Emu Mountain (73m), which was adjacent to the coast. From this vantage point, visual observations were possible out to 15 km from the survey point. For further information on the study site set-up and calibration of the acoustic array see Noad et al., 2004; Dunlop et al., 2007 and 2008; Smith et al., 2008. Visual data collection involved two platforms of observation; '*ad lib* sampling' and 'focal follow'.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Visual platform of observation

Land-based behavioural observations were collected daily (7am to 5pm, weather permitting). A theodolite (Leica TM 1100) was connected to a notebook computer running Cyclopes software (E. Kniest, Univ. Newcastle, Australia) and used to track and observe passing whales. In this study, the sample unit was a group of whales, defined as those whales surfacing within 4 body lengths (about 50 m) of each other. When whales are travelling, their surfacing intervals are usually several hundred metres apart, much larger than the spacing of the whales from each other within a group and far less than the spacing between groups. Group sizes usually comprise one to three individuals. Cyclopes records the positions of whales from the theodolite elevation and azimuth in real time. Fixes were annotated with observed behaviours and group compositions out to a 10 km limit. Two observers with binoculars were responsible for keeping track of all visible groups in the area as *'ad lib'* observations (including the target group during an experiment) and directing the theodolite operator to groups to be fixed. Data from the visual observers included bearing and distance from Emu Mountain, group composition, group behaviours (blow, breach, pectoral flipper slap, tail slap, splitting apart of a group, joining together of two groups, no blow rise or surfacing, peduncle slap, inverted tail slap, inverted pectoral flipper slap and head lunge being the majority observed) and direction of travel. These were recorded by the Cyclopes operator (as 'additional observations' made using binocular bearing and reticule readings).

Focal follow platform of observation

The focal follow method of observation was introduced in the 2008 experiment. A dedicated focal follow team was situated slightly apart from the *ad lib* sampling team to avoid confusion, but within audible range to allow some communication between teams. This team consisted of a theodolite operator and a Cyclopes data recorder. Once a suitable group was targeted for an experiment by base station, the focal follow team concentrated only on this group for the duration of the experiment where all visible behaviours were recorded.

Digital recording tag platform of observation

A Dtag (non-invasive, digital acoustic recording tags with depth and orientation sensors, with acoustic sampling rate 64kHz and sensor sampling rate 5Hz, Johnson and Tyack, 2003) was deployed onto a mother within a female-calf group during one of the 'social sound' experiments. The tag was attached to the back of the whale as she surfaced in front of a specially-equipped boat, using a long pole attached to the bow. The tag was attached by suction cups and pre-programmed to detach after four hours. It contained a hydrophone and three-axis accelerometers and magnetometers to measure pitch, roll and heading (Johnson and Tyack, 2003). An estimated 3D dead-reckoned track including the dive profile can be derived using Dtag data and an estimate of travel speed (Miller et al., 2009). The Dtag hydrophone provided a high-quality recording of the sound field at the whale.

Exposure stimuli

A J11 acoustic projector was used as the underwater loudspeaker. It was suspended 10 m below a small boat which was allowed to drift. A hydrophone was suspended to the same depth from the other end of the boat to measure the J11 source level (by correcting for the propagation loss between the J11 and the hydrophone assuming spherical spreading). The 'social sound' stimulus consisted of a 20 min sequence of social vocalisations compiled from a variety of social sounds recorded using a Dtag deployed previously onto a female-calf-escort group passing through the site. The escort was probably a male as groups with two or more adults and a calf generally consist of an adult female, calf and one or more male escorts (Baker and Herman 1984; Tyack and Whitehead, 1983). A collection of different social sounds was spliced together to make up a recording of 204 s duration. This was repeated to make up the 20 min stimulus. We decided to

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

use only one recording of social sounds as the goal of the study was to look for difference in response to a recording of ‘natural’ conspecific sounds (which, based on previous work, we assumed would produce a reaction) compared to an unnatural ‘tone’ sound (following the design of Mobley et al., 1988). We assumed that using different recordings of social sounds would produce highly variable reactions (dependent on the sound types as well as the social context of the recorded group) and therefore, to reduce the potential variability in the reaction, used only one recording. However, using only one social sound stimulus does limit any conclusions that can be made about the observed response to social sounds as well as the function of these sounds. To negate external validity issues and make more generalised conclusions on any differences in response to ‘tones’ versus ‘social sounds’, it would have been better to repeat the study using a different set of social sounds. The tone stimulus consisted of sequence with a tone swept in frequency from 2 kHz to 2.1 kHz over a period of 1.5 s, repeated every 8 s for 20 min. Source levels varied from 148 to 153 dB re 1 µPa at 1 m rms for both stimuli (similar to source levels of humpback whale social vocalizations, unpublished data). Stimuli were recorded on a CD and played through an amplifier into the J11. Also recorded on a CD was a silent control consisting of a 20 min recording with no signal input.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Experimental design

This experiment followed a typical ‘BDA’ (before, during, after) design. The ‘B’ period consisted of a pre-exposure (stimulus off) control, the ‘D’ period was a period in which the stimulus was turned on, followed by a post-exposure (stimulus off) ‘A’ period. Each period lasted for 20 minutes. Exposure treatments were one of three; a silent control, a recording of conspecific social vocalisations or the artificially-generated 2 kHz tone. To increase the sample size of the control treatment, groups which migrated within 2 km of a moored vessel (the research boat or a similar sized vessel) were also included in this ‘control’ category. Therefore not all ‘control’ treatments involved a J11 being deployed in the water playing ‘silence’.

Baseline groups migrating through the study area, selected randomly, were focally followed for at least one hour during times when no experiments were underway. We also selected a number of *ad lib* sampled, baseline groups for analysis based on the following

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

selection criteria: 1) they had to be visually tracked within the study site for at least one hour and 2) they did not move within 2 km of a stationary vessel during the hour specified for analysis. This comprised the ‘baseline’ dataset.

Movement response variables

Movement response variables (measures of how the group travelled through the study site) such as speed, course and distance travelled between each surfacing were calculated by examining the difference in position between each theodolite fix. A measure of how erratic the group course was determined by the difference in course between successive fixes. The total distance travelled within each period (taking into account all changes in course) was calculated by summing all distances between consecutive surfacing events for that period.

Only one theodolite fix was chosen (usually the first fix on the group within a surface interval after a deep dive) to represent each group surfacing. Generally animals within each group were less than 50 m apart therefore this tracking method provided the best representation of group movement through the study area. If surfacing events were missed within experimental periods (in the *ad* lib sampling dataset), the assumption was made that groups travelled in a straight line and at constant speed between the two consecutive surfacing events. The mean of all measurements of course travelled (magnetic bearing), variation in course travelled, and speed were calculated for each 20 minute experimental period. The ‘course made good’ for each period was estimated by using two fixes: the one at the start of the experimental (BDA) period and the one at the end, and calculated as the bearing of the second fix relative to the first.

Behavioural response variables

The behavioural response variables consisted of measures of diving and surface behaviour. Dive profile incorporates “surfacing” dives (the short and shallow dives that occur during respiration bouts) and “long dives” in which the group disappear for a longer period of time. A “long dive” is defined as the time from when the last group member disappears to when the first group member re-appears and the “surface interval” is defined as the time spent on surface between long dives which incorporated all “surfacing dives”. Discriminating between “long” and “surfacing” dives can be problematic. Typical humpback whale dive pattern tends to be a number of short respiration dives followed by a longer dive (usually lasting three to five minutes). Focal follow data were used to differentiate between shallow respiration dives and long dives as the majority of surface behaviours from each target group should have been recorded and timing of these events should be relatively accurate. The time between each successive sighting (dive time) was measured within each group and the log transformed time (due to non-normality) was plotted as a histogram. This gave a bimodal histogram; one peak corresponding to peak respiration dive times and one peak corresponding to long dive peak times. We used a probability density function in the histogram as a guide to determine the two peaks in the dive time dataset as well as an appropriate cut-off time between respiration and long dives (estimated as the trough between the two peaks). This provided a separation value of 60 seconds. Dive times of less than 60 seconds were designated ‘short’ respiration dives and dive times longer than 60 seconds were designated ‘long dives’. The peak respiration dive time was found to be ten seconds (times ranged from 2 to 58 seconds). The peak long dive time was found to be three minutes (ranging from 60 seconds to 18 minutes). Inspection of the final long dive dataset showed that 18 minutes was an outlier (it may have been two long dives), therefore we omitted this point, leaving the range of long dive times to be between 60 seconds and 11 minutes. The number of long dives and surface intervals (which included all respiration dives) and the mean durations of these dive profile behaviours were calculated for each experimental period.

Surface intervals were classified as either ‘blow only’ (no animal within the group was surface active during the surface interval) or ‘surface active’ (one or more animals within the group were surface active during the surface interval, in other words, breaching, pectoral slapping or tail slapping behaviour was observed). The number of each type of surface interval was counted for each experimental period.

Social variables

Whale groups were divided into five different categories based on typical composition of groups observed during the southern migration (Table 1); lone animals (which may or may not have

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

been singing during the experiment), female-calf groups, adult pairs, female-calf-escort groups (the escort may or may not have been singing during the experiment) and groups with more than three adults (female-calf-multiple escorts or groups of three or four adults). However, due to the small sample size of each cohort, groups were divided into ‘female and calf’ groups (containing a female and no adult male) and ‘lone’ (many of them were singers and therefore males) and multiple (all other cohorts). It is likely that the presence of an escort, or number of escorts, in a group, including a mature female, will have a significant effect on group behaviour compared to a group containing only a mature female, with or without a calf. Female associations are thought to be rare (Brown and Corkeron, 1995; Clapham, 2000) and humpback interactions involving groups with two or more adults and a calf generally consisted of an adult female, calf and one or more male escorts (Baker and Herman 1984; Tyack and Whitehead, 1983). The group composition of all other groups in the study area and the distance of each group from the target group were noted throughout the experiment. For this analysis, only the presence of the closest group (the ‘nearest neighbour’), the mean distance of the nearest neighbour from the target group during each experimental period, and the mean distance of the nearest singer from the target group during each experimental period (estimated using acoustic positions overlaid on top of visual positions) were considered as social factors. We also noted if a group joined the target group or the target group split into two smaller groups within each experimental period.

Insert table 1 here

Table 1. The sample size of different measured cohorts for the two visual platforms of observation

The environment

Wind speed was measured using a weather station placed on the roof of the base station. The mean wind speed was calculated for each experimental period.

In the study area, sounds from singing whales were frequent components of the underwater noise environment, though small recreational vessels were often audible as they traversed the area. During this experiment, the majority of the samples had little interference

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

from vessel noise and therefore background noise levels (without singers) was mainly typical ambient noise (Cato, 1997), mostly due to noise from sea surface motion (wind-dependent noise) and snapping shrimps. Traffic noise, the noise from distant shipping, is significant further off shore, but the shallow water approaches to the site would have limited this contribution. In many cases, noise measurements could be made without significant contribution from singing humpback whales. When song made significant contributions to the noise, the noise in the absence of song was estimated from the periods in between identifiable song units. To do this, a recording was displayed as a wave form (Adobe Audition) and song units were deleted leaving only the time periods between song units. A 20 second noise sample was obtained in this way. Song units were usually separated by 1 – 3 s and the song fades out as the singer comes to the surface to breathe. This may have contained undetectable song units, but these would not have made a significant contribution to the estimate of wind-dependent background noise levels. During exposure, the noise was estimated in the same way by deleting the periods when the stimuli were present. A 20 second noise sample was taken from each hydrophone in the array every 10 minutes, starting 10 minutes before the start of the experiment and ending 10 minutes after the finish of the experiment. The noise in each 20 second sample was measured in one-third octave bands levels in the range of and the system calibration applied to obtain levels in dB re 1μPa. One-third octave bands represent the logarithmic increase in frequency range of auditory filters in the mammalian ear (Fletcher, 1940) and in humpback vocalisations most sound energy of the fundamental frequency is contained within a one-third octave band, making this an appropriate filter. The total background noise level was calculated by summing the mean square pressure for each one-third octave band for the frequency hand of interest (as indicated below) and converting this to total broadband noise level (dB re 1μPa). Mean broadband noise levels for each experimental period were then calculated from all samples taken from all hydrophones.

Background noise levels (excluding contributions from singers) at the array were assumed to be similar to those at each humpback whale group, since it was predominantly wind-dependent noise and wind speed was generally uniform throughout the study site (snapping shrimp noise did not contribute significantly in the frequency band of interest). This was not the case for noise from nearby singers, which was dependent on the distance of the singer from the receiver. Therefore, analysed groups were also categorised according to the social environment: ‘none’ (no audible singers present); ‘close singer proximity’ (the nearest singer was within 2 km of the group, or became part of the group, such as a mother and calf being joined by a singing escort), ‘medium singer proximity’ (the nearest singer was between 2 and 5 km from the target group) and ‘far singer proximity’ (the nearest singer was more than 5 km from the group).

Received levels and signal to noise ratios

All received levels of each stimulus were measured in one-third octave bands from recordings made on the fixed array using SpectraPLUS 5.0 (Sound Technology Inc.). Three ‘tones’ were selected for measurement in first 10 minutes and 3 ‘tones’ in the second 10 minutes of exposure. The sound pressure levels at the array for each ‘tone’ sound were measured (dB *re*. 1µPa) in the 2000 Hz one-third octave. For the social sounds, one of the highest level sound types was chosen for measurement. Six samples were measured in one-third octave bands from 200 to 400 Hz (centre frequencies) which contained most of the energy.

These measured levels have contributions from both the signal and the background noise. In order to obtain a true measurement of the received signal level, the contribution of background was removed. The noise levels measured just before and after exposure were used as the noise sample to estimate the noise level *NL*. To obtain the received signal level *RL*, the levels measured during the stimulus and noise levels were converted to mean square pressures and the noise mean square pressure was subtracted from the measured mean square pressure during the stimulus for each one-third octave band

$$< p_s^2 > = < p_m^2 > - < p_n^2 >$$
$$RL = 10\log(< p_s^2 >)$$
$$NL = 10\log(< p_n^2 >)$$

where $< p_s^2 >$ is the mean square pressure for the signal, $< p_m^2 >$ is the measure mean square pressure during exposure and $< p_n^2 >$ is the measured mean square pressure of the noise. The resulting one-third octave band signal mean square pressures were summed and then converted to decibels (10log (summed mean square pressure)) to give the received signal level. Ambient noise levels at a whale groups could be assumed to be similar to those at the array, since the noise is predominantly wind dependent. Signal levels received by a whale group would, in

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

general, differ from those at the array because of the differences in distance from the source and thus transmission loss. The signal level received by the group was therefore determined by correcting for the difference in transmission loss between the source to the array and the source to the group.

Transmission loss was measured using the noise generated by a noisy boat as the source. The boat conducted runs along lines radiating from the array, from distances of 100 m out to about 10 km from the array. Regression lines were fitted to the data as a function of the logarithm of the distance. The results were in the form of relative loss over the distance of measurement of the form $\mathrm{TL} = a + b\log(x)$ where b is the slope of the regression line, x is distance and a is a constant. The received level at the group could then be determined from the received level at the array by $RL_\mathrm{g} = RL_\mathrm{a} + b\log(x_\mathrm{a} - x_\mathrm{g})$, where RL_g and RL_a are the received levels at the group and the array respectively, and x_g and x_a are the distances from the playback source of the group and the array respectively. For most frequencies, b varied with distance but could be well approximated by two values, one applying to distances less than and the other greater than a cross over value.

For received levels at the whale group that are close to noise levels, signal audibility or detectability is likely to change with varying background noise levels as well as absolute received level, so may influence the response. We therefore included signal to noise ratio at the group as an indicator of audibility as an exposure metric, estimated as

$$SNR_g = SL_g - NL_g$$

where subscript g refers to the value at the group, and SL_g is the signal level at the group. Measurements for each term were made in the following frequency bands: 2 kHz one-third octave (1782 – 2245 Hz) for tones and over the 200 – 400 Hz one-third octaves(177 – 446 Hz) for the social sounds. Note that it is possible for estimates of SNR_g to be negative if the distance from the source to the group is significantly larger than from the source to the array.

The received level of each stimulus at whale groups varied over a range of 40 dB while the ambient noise varied over a range of 30 dB. It is possible that some of the lower received levels were masked by the ambient noise background and thus not heard by the whales. Masked thresholds of audibility have not been measured for humpback or any other species of baleen

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

whale. However, they have been measured for a range of terrestrial and marine mammal taxa and there is a broad consistency of results. While the extent to which this information can be applied to humpback whales is limited, it does at least give an indication of where the signals may be below the masking threshold and thus inaudible.

One measure of masked threshold is the critical ratio which is the difference (in decibels) between the level of a tone at the threshold of audibility and the spectrum level of white masking noise at the same frequency (Richardson et al., 1995). This is close to the situation in the masking of the playback tones by ambient noise. Masking of a tone is considered to be cause by a limited bandwidth of the noise, typically less than 20% of the tone frequency at 2 kHz, and over this band the ambient noise is a reasonable approximation to white noise. Critical ratios measurements for various species are summarised by Richardson et al. (1995) and Southall et al. (2007). The value at 2 kHz ranges from 19 to 26 dB across several species of pinnipeds and is 19 dB for the beluga, 20 dB for humans and 25 dB for cats. These results provide the best information we have to infer where the playback of tones might be masked by the ambient noise. Our measurements of SNR for the tones used the noise level in the one-third octave band at 2,000 Hz, i.e. over the band 1782 – 2245 Hz, a bandwidth of 450 Hz. Noise levels in this band will be 10 log (463) = 26.7 dB higher than the spectrum level, so that *SNR* using the one-third octave band for noise will be 26.7 dB lower than using the noise spectrum level. The range of critical ratios of 19 – 26 dB are thus equivalent to *SNR*s using the one-third octave band for noise of – 7.7 to –0.7 dB, an average of –4.2 dB. The analysis of the 'tones' experiment was therefore conducted using a subset of the data limited to $SNR \geq -4$ dB, to exclude data that might have been inaudible, as well as using the full data set.

Critical ratios are generally measured for tonal signals and there does not appear to be measurements applicable to signals like the social sounds. The social sound type chosen for the analysis has most energy extending across three adjacent one-third octaves (centre frequencies 200 – 400 Hz, i.e. from 177 to 446 Hz) and we measured the *SNR* for both the signal and the noise in this band. If the masking frequency band is wider than the signal band, as it is for tones, the threshold of audibility would occur for $SNR > 0$ dB (signal and noise measured in the same band). However, the social sound used is harmonic, and for any harmonic, the masking bandwidth may be significantly less than this, i.e. closer to the masking band for a tone. This would suggest a threshold of audibility significantly at a *SNR* of significantly less than 0 dB for the way we measured *SNR*. In the analysis, a subset of data which excluded *SNR* < 0 dB at the start of the 'during' phase was used to exclude playback that might have been inaudible. As it happened, the highest *SNR* experienced by the whale groups during exposure exceeded 6 dB (as groups approached the source vessel) for all included groups, so it seems unlikely that any in this reduced dataset were not audible, at least for some part of the exposure.

Statistical Analysis

All analyses were generated using the statistical software package 'R' (R Foundation for Statistical Computing). To test for sampling bias between BDA periods, in other words to test if there was a more concentrated effort in the 'D' period, the (normalised) mean number of observations between experimental periods was compared in both the 'focal follow' and '*ad lib*' data. No sampling bias was apparent. A measure of group visibility was compared between experimental periods to test there was any bias in group sightability due to increased sighting effort, increased time spent on the surface or increase in surface active behaviours making the group more visible and less likely to be missed. The measure of group visibility used was the total amount of time per BDA period that groups were sighted on the surface (or in a shallow surfacing dive) expressed as a percentage of the total time of each experimental period. These percentages were compared between periods and no significant difference was found. As a result, all observations were used in the dataset.

The mean (+ standard deviation) of each response variable (course travelled, change in course travelled, speed travelled, number of deep dives, number of surface intervals, length of deep dives, length of surface interval number of 'blow only' surface intervals, number of 'surface active' surface intervals, course-made-good and distance travelled) was calculated per experimental period for each humpback whale group. Linear mixed effects models were fitted to each response variable which included the random effect of 'group' (and associated variance). Standard statistical models assume independence of errors, but when measurements are taken from the same group, they are correlated. Mixed-effects models account for interdependence in multiple observations within individuals as they assume the data within groups (in this case, each

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

vocalising group) are dependent among the observations and model the covariance structure introduced by grouping the data. The included random effect estimates the distribution of the means as a standard deviation of the differences of the factor-level means around an overall mean, instead of estimating a mean for every single factor level. To test the effect of stimulus exposure on behavioural measures, linear mixed-effects models (using the lme4 package, Bates et al., 2011) were used which included stimulus type, experimental period, environmental and social variables and measures of received level and signal-to-noise. ‘Group ID’ was included as a random factor. Models including different terms (null and predictor variables) were compared using Akaike Information Criterion scores and checked for significant ($p<0.05$) improvement using the maximum likelihood ratio (LR) test, where the probability distribution of the test statistic is a chi-squared distribution and the degrees of freedom equals $df_1 – df_2$ (where df_1 and df_2 are the degrees of freedom for the two models being compared). Mixed fixed effects models can be problematic as the distribution of the fixed effects is uncertain under the null hypothesis and the denominator degrees of freedom for tests are difficult to determine (Bates, personal communication). Therefore p-values were generated using the Markov Chain Monte Carlo (MCMC) method using the ‘language R’ package. Residuals of each model were checked for homoscedasticity and errors were checked for normality. Within model *t* values with associated p-values are also reported for specific comparisons.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

RESULTS

Sample size

In 2008, 15 experiments were carried out; 8 using the social sound stimulus, 6 using the tones stimulus and 1 silent control. A further 6 groups were used as controls. All focally followed groups were from the 2008 experiment and only 1 group was focally followed per experiment. In 2004, 16 experiments were carried out; 7 using social sounds, 7 using the tones stimulus and 2 using a silent control. All groups in 2004 were sampled ‘*ad lib*’ (as much data on each group in the area was collected without focussing on one specific group) and multiple groups were sampled during each experiment. A further 19 groups were selected as baseline groups from the two years. *Ad lib* sampled groups were also used in 2008. Table 2 presents the sample size of groups used for the analysis combining both 2004 and 2008 datasets.

Insert table 2 here

The experiment was carried out on southerly migrating groups (in a population of over 10,000 animals) therefore it is highly unlikely that any group was repeatedly sampled. If the group split into two separate groups (n = 8), only one of those groups was used (the one which appeared first after the split).

All focal follow samples can be considered independent as only one group was focally followed during each experiment. All baseline samples were also independent (one sample per day). Of the *ad lib* sampled groups exposed to either social sounds or tones, 28 groups were multiple samples, in other words, during any exposure experiment, up to three groups may have been used for the analysis. In 2008, one of these groups would have been also focally followed. In 2004, all groups were *ad lib* sampled. If groups do not interact with each other in such a way that the response to the stimulus is influenced by this interaction, then they can be considered as independent samples (Miller et al., 2009). We minimised the potential for non-independent sampling by ensuring the following criteria was met. No groups that were simultaneously used in the analysis interacted with each other (in other words, joined together) and none of these groups came within 3 km of each other (average distance apart was 5755 m, range 3000 – 10,000 m). We used a 3 km limit as the most likely interaction between groups would have been mediated acoustically and it is difficult to hear social sounds on the array from groups beyond 3 km. This minimises the risk that the groups were somehow influencing each other’s behaviour. To further check this, we looked for social sounds on the acoustic recordings made during each trial and found that no sampled group that was also vocalising was within 4 km of any other simultaneously sampled group.

We also accounted for the effect of the ‘nearest neighbour’ group (not usually another sampled group) to determine if nearby groups had any influence on the behavioural response parameters. While socially vocalising groups are unlikely to be heard more than a few kilometres away, singing whales are audible over distances of tens of kilometres and therefore could potentially affect the behaviour of any group within audible range. In the analysis we also

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

accounted for the presence of the nearest singing whale as a fixed effect (assuming the nearest singer is more likely to have an influence on the behaviour of the group compared to more distant singers).

Dtag

Only one experiment was carried out using a Dtag. The tagged animal (the female from a female-calf group) changed dive behaviour to shorter, shallower dives during the time the social sound stimulus was played and did not return to pre-exposure dive behaviour after exposure (Fig. 1). The animal also changed direction, from consistently travelling at a mean of 225 degrees (south westerly direction) to head directly west (inshore), then north. After the experiment had finished, the group slowly returned to a southerly course. This group was also tracked from the visual station (though was lost during exposure, probably due to the change in dive behaviour resulting in the animals becoming very difficult to track). From the dead-reckoned track, the distance from the source vessel at the start of exposure was estimated (using received levels measured at the array and then estimated at the group) to be 880 m (signal level *RL* of 101 dB re 1 µPa and *SNR* of 8 dB) and the distance from the vessel when the group initially changed course was 660 m (signal level *RL* of 105 dB re 1 µPa and *SNR* of 13 dB).

Insert figure 1 here

Visual observation data

A total of 15 groups were both *ad lib* sampled and focally followed at the same time from the two different platforms of observation. Each response measurement for each platform of observation was averaged over each experimental period. A mixed effect model was used to test if there was any difference in the any of the response measurements between the two platforms of observation where ‘platform’ (*ad lib* or focal follow) was included as a fixed effect and group ID as the random effect.

Insert table 3 here

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Movement response variables (course travelled, variation in course travelled) and a two of the behavioural response variables; the number of surface intervals and number of ‘blow only’ surface intervals were comparable between the two visual survey platforms. However, behavioural measures such as ‘long’ dive times and mean surface interval times were found to be significantly different. This suggests that all of the surface intervals are being captured in both the focal follow data and *ad lib* sampling data, however the timing of behaviours such as the long dive times and surface interval times were significantly different due to the *ad lib* sampling team missing a number of group surfacing behaviours (as shown by the difference in the number of observations per experimental period for each of the data platforms).

To increase the experimental power (by increasing the sample size) and allow the incorporation of other factors into the analysis model, we pooled the data from both platforms of observation (using focal data from groups which were both *ad lib* sampled and focally followed) when testing all movement variables and when testing numbers of behavioural events, but not when testing the timing of events. Only five groups were exposed to ‘silence’ therefore we pooled these data with baseline groups (after first comparing response variables between ‘non-exposed’ and ‘silent’ groups and finding no significant difference). These groups will hereafter be referred to as ‘baseline’ groups.

The response to stimulus and experimental period

The following analysis includes only groups in which we assumed the stimulus was audible at some stage during exposure (reduced dataset) and all ‘baseline’ groups.

The course travelled by groups (n = 53 groups) was dependent on the stimulus type combined with the experimental period (LR $\chi^2_8 = 31.7$, $p = 0.0002$) and results from this model suggest that groups exposed to ‘tones’ generally travelled on a more south-easterly (offshore) course during exposure (change in course estimated at -20° relative to the before phase of baseline groups, SE = 13.6, t = -4.6, p = 0.001) and after exposure (change in course estimated at -12°, SE = 13.6, t = -4.8, p = 0.0006 relative to baseline groups) compared to baseline groups (whose course was estimated at 177° (SE = 9.0)). Groups exposed to the ‘social sounds’ recording and baseline groups tended to migrate in a south-south-west direction, following the

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

coastline (there was no significant difference in travel direction). However, some groups visibly changed direction when exposed to the 'social sounds' stimulus, though usually returned to their previous course at some point during exposure. Looking just at the focal follows ($n = 8$), some groups obviously changed course and approached the boat to within 100 m (one single animal, one female-calf-escort group and one pair) whereas other groups (for example, the tagged female from the female-calf group) moved inshore and away from the vessel at some stage during the playback of social sounds. In one instance, a singer stopped singing and moved away from the vessel, whereas, in two instances, a single animal split from a group and started singing in close proximity to the vessel. Therefore we found a highly variable but not prolonged response in terms of the change in course travelled in groups exposed to our recording of 'social sounds', whereas response to 'tones' was a consistent and prolonged change in course to a more offshore direction.

The (normalised) number of surface intervals per 20 minutes was found to be significantly dependent on the experimental period combined with the stimulus type (LR $\chi^2_8 = 32.2$, $p < 0.0001$; reduced 'audible' dataset) as was mean 'long' dive time (LR $\chi^2_8 = 32.6$, $p < 0.0001$; focal follow dataset). Fig. 2 illustrates the changes in dive time (focally followed groups; $n = 20$) and number of surface intervals per experimental period ($n = 53$) during the experiment for baseline and exposed groups. Groups exposed to 'tones' displayed a greater number (estimated at 1.5 surface intervals per experimental period) of surface intervals during exposure ($t = 3.7$, $p = 0.0001$) compared to baseline groups (which surfaced about 3 to 4 times per experimental period) and a decrease (estimated at 106 seconds) in dive time ($t = -2.2$, $p = 0.03$) compared to baseline groups. The number of 'blow only' number of 'surface active' surface intervals and the length of the surface interval were not found to be significant response variables.

The effect of environmental and social variables

Environmental variables such as 'wind speed' or 'background noise levels' and social variables, such as the 'number of groups in the study area', 'social composition of the nearest neighbour' or

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

‘distance of the closest neighbour’, were not significant predictor variables in any response model.

We added in the ‘social composition of the groups’ (lone animals and lone singing whales were categorised together as ‘lone animals’, female-calf pairs formed their own social category and adult pairs, female-calf-escort(s) and groups with more than two adults were categorised together as ‘multiple adult groups’) to the course travelled response model (which included the term stimulus only) and found a significant (LR χ^2_6 = 17.7, p =0.006) improvement in this model. Female-calf groups, in response to ‘social sounds’, tended to take a much more westerly (inshore) course compared to ‘multiple adult’ groups (t = -3.1, p = 0.003). The response to ‘tones’ in terms of course travelled was similar within each group social composition for all datasets.

There was also a significant (LR χ^2_6 = 28.7, p =0.0001) improvement in the number of surface intervals response model. This was due to the differences in dive behaviour between the social categories. ‘Lone animals’ in general tended to surface significantly less often compared to female-calf pairs (t = -2.8, p = 0.03) and ‘multiple adult’ groups (t = -5.1, p = 0.0001). However, although most groups responded to ‘tones’ by increasing the number of surface intervals, the response to ‘social sounds’ was again highly variable. Some groups increased the number of surface intervals and others decreased the number of surface intervals during exposure, but no significant trend with social category was found. However, sample size for each social category was quite small.

The effect of source proximity, received signal levels and received *SNR* variables

The following analysis includes only groups exposed to either stimulus (n = 37) for the pooled ‘*ad lib*’ plus focal-follow dataset (testing course and number of surface intervals) and 14 for the focally followed groups (testing ‘long’ dive time) to test the effect of proximity of the group to source, received signal level (*RL*) and received signal-to-noise level (*SNR*) at the start of exposure on each response variable. To test which of the exposure metrics (proximity to source, *RL* or *SNR*) best predicted the response we compared four different models for each response variable within the two different datasets; the full dataset (including probably ‘inaudible’ experiments) and the ‘audible’ dataset (including only those which we assume are audible as defined by the previous criteria). The following four models were compared:

1. Model including stimulus and experimental period only,
2. Model including stimulus, experimental period and proximity,
3. Model including stimulus, experimental period and *RL* and,
4. Model including stimulus, experimental period and *SNR* as predictors.

Insert table 4 here

The inclusion of *SNR* as the exposure metric significantly improved the response model for course travelled, though only in the full dataset. The best exposure metric to predict the response in terms of the number of surface intervals was both *RL* or *SNR* (full dataset) and *SNR* in the ‘audible’ dataset. For long dive time (using only focal follow data which we assume all are audible) the best exposure metric was proximity to the source (Table 4).

Groups changed their course to a more easterly direction during (t = -2.2, p = 0.02) and after (t = - 2.7, p = 0.009) exposure to ‘tones’ as the received *SNR* of the signal increased at the start of exposure (Fig. 3). *SNRs* ranged from -22 to 15 dB at the start of exposure and the proximity to the source at the start of exposure ranged from 300 m to 8.8 km in these groups. Though we suspect the ‘tones’ were only audible from about -8 dB (at a distance of about 3.5 km depending on the background noise). The received signal levels of the ‘social sounds’ stimulus ranged from 72 to 98 dB re 1µPa, *SNR*s ranged from -23 to 21 dB and the proximity of the group at the start of exposure ranged from 440 m to 8 km but groups did not respond to this stimulus in terms of a consistent change in course and therefore it was not possible to assess the effect of any exposure metrics.

Groups, when exposed to ‘tones’ also increased the number of surface intervals as the received *SNR* increased at the start of exposure (t = 2.1, p = 0.02, Fig. 3) and the *SNR* was found to be the ‘best’ exposure metric for predicting this response for all datasets. An increase in the *SNR* at the start of exposure also resulted in a decreased number of surfacings post exposure (t = -2.2, p = 0.03). Groups tended to surface less often during exposure to ‘social sounds’ compared to groups exposed to ‘tones’, however, there was no real trend with *SNR* in these groups

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

(probably due to the variation in reaction). In other words, the relationship between the response variable and the *SNR* at the start of exposure was found only in groups exposed to 'tones'.

The long dive time response was significantly related all three exposure metrics though the proximity of the group at the start of exposure was the best predictor of the response (Table 4). However, these focally followed groups were always within 2 km from the source at the beginning of the exposure phase (proximity ranged from 300 m to 2 km, *RL* ranged from 84 to 112 dB re 1µPa and *SNR* ranged from 2 to 14 dB). The proximity to the source vessel had an effect in the post-exposure phase, where groups exposed to 'tones' displayed a decrease in long dive time with decreased proximity to the source and groups exposed to 'social sounds' displayed an increase in long dive time with decreased proximity to the source ($t = 1.8$, $p = 0.05$).

Insert figure 3 here

DISCUSSION

Out of all tested response variables, three (course travelled, the number of surface intervals and long dive times) were found to change significantly in response to three exposure metrics: proximity, the received signal level (*RL*) and the signal to noise ratio (*SNR*) at the group. An easterly change in course (away from the coast) was found to occur during and after exposure in response to 'tones'. These groups also tended to spend more time close to the surface (by increasing the number of surface intervals and decreasing dive time) during exposure. The magnitude of the change in course and dive behaviour was related to the proximity, *RL* and the *SNR* of the stimulus at the start of exposure. Groups exposed to our recording of social sounds did not significantly change their direction of migration, though we did find short-term changes in travel direction. Different social groups reacted quite differently to this stimulus; female-calf groups tended to move inshore and spend more time near the surface. Other social groups approached the source vessel but returned to their original travel direction at some point during exposure. This paper presents evidence that migrating humpback whales differ in their behavioural response when presented with a recording of conspecific social sounds compared to artificial tones, and this change in behaviour was influenced by other factors: the social group, the proximity of the group to the source vessel and the initial 'dose' (as measured by the start signal-to-noise ratio and received signal level).

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

A change of course was most evident in groups exposed to ‘tones’, where groups moved away from the source vessel and offshore at some point during exposure, indicative of an avoidance reaction to this stimulus. In comparison, many groups (mainly those thought to contain a male because one was a singer or an escort with a female and calf) exposed to ‘social’ sounds first approached the source vessel, then at some point resumed their previous course, or continued along their path towards the source vessel during exposure. Both previous behavioural response studies in humpback whales using conspecific social sounds found that the social composition of the group was an important factor in determining the response. Tyack (1983) found that singing males stopped singing when either song or social sounds were played and the majority of them ‘charged’ the boat when exposed to social sounds. However, females with calves and large groups tended to move away from the boat during exposure to these sounds. Mobley (et al., 1988) found rapid approach responses in singletons and adult pairs but no approaches by females with a calf. Although our sample size (with focally followed groups) was small, we found similar results, with some single animals and adult pairs approaching the boat, whilst some females with calves evidently changed course to avoid the source vessel (though they tended to move inshore). This avoidance reaction (in terms of a change in the direction of travel during exposure) was very clear in the single tagged group. However, this study only used one recording of social sounds and therefore inferring the function of these sounds based on the observed behavioural reactions goes beyond the scope of the study. Although the sample size of this study, in terms of determining the social effects, was limited, it demonstrates the complexity of behavioural responses to stimuli and the need to measure as many other factors as possible (and generate a large sample size) in order to tease out such complex interactions. It would be beneficial to repeat the study with a different set of social sounds to negate external validity issues with only using one stimulus (allowing us to make more generalised conclusions on the difference in response to ‘tones’ compared to ‘social sounds’). These experiments could also be targeted towards testing the function of specific sounds by using a number of different recordings from various cohorts to determine if there are consistent avoidance and attraction responses to each combination of vocal signals.

In this study we found diving and surfacing behaviour also significantly changed with exposure to both test stimuli. Previous studies assessing the behavioural response of humpbacks to an M-sequence sound (Frankel and Clark, 1998) and a recording of a full scale Acoustic

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Thermometry Ocean Climate (ATOC) sound source signal (Frankel and Clark, 2003) found responses such as increases in time between surfacing events and a greater distance travelled underwater. In other words, they found exposed humpback whales tended to spend more time underwater and travel further compared to baseline groups. In this study we found that groups consistently increased the number of surface intervals (and consequently decreased the dive time and therefore time spent at depth) in response to 'tones'. This may indicate an avoidance reaction to our signal, but that avoidance reaction differs to that found by Frankel and Clark (2003). We found that female-calf groups tended to respond to 'social sounds' in a similar way and their change in dive behaviour to more frequent yet brief surfacing events. This may also be a way of females with calves avoiding what was perceived by them as a nearby group that might contain a male.

The Frankel and Clark experiments included group composition, the presence of nearby vessels and the received level as additional predictor variables. Most cetacean behavioural response studies to date have considered only the received level (Southall et al., 2007). However, the relative level of the signal compared with the background noise (*SNR*), or the signal excess above masked hearing thresholds, may be significant predictors of behavioural response, and may under certain conditions (such as when the receiver is at greater distances from the source and received levels are close to background noise) be a better predictor than received sound pressure level. To that end we found the *SNR* to be a better predictor of behaviour change response (in terms of course travelled) than received level and proximity to the source when using the full dataset (where groups ranged from 300 m to 8.8 km from the source). The change in dive behaviour (measured by the increase in the number of surface intervals) in response to 'tones' was also highly related to the *SNR* at the start of exposure. We could only measure long dive time using focal follow data (where groups were within 2 km from the source) and, using this dataset, did not find that *SNR* was the best predictor term. Rather proximity to the source was. Results of behavioural response experiments are often used to inform management of the effects of noise on marine mammals. This study shows that care must be taken when choosing which exposure metric (proximity to the source, received level or *SNR*) to use when predicting dose-response relationships as results could be highly dependent on the range of data chosen as well as the response variable.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The relationship between signal to noise ratios and masked auditory detection thresholds of signals against noise is complex. It seems likely that most experiments in this study would have been audible, but, given the variability of ocean noise, it is possible that some of the full data set may not have been and some only intermittently so. However, the subset of data should have excluded most samples where the experiment was inaudible and so long as a whale can hear a sound, there is the potential for a behavioural response. Higher signal to noise ratios might be more likely to attract a listener's attention and it is possible that *SNR* is used to judge signal level and thus proximity of the source. Hence, it might be expected that *SNR* would be an important exposure metric to dictate the response. On the other hand, the dose-response relationship may be lost when using only high *SNR* experiments. Therefore, including experiments with low received signal levels may help to determine the threshold of response and provide some clue as to the auditory sensitivity of these animals. Whether responses to low level signals have longer term significance is, of course, a different question.

This study is one of the more comprehensive behavioural response studies that have been carried out on a large whale species. Sources of pseudoreplication were considered (a limitation of the study being that only one recording of 'social sounds'). We used two different stimuli and applied a statistical analysis that accounts for individual variation as well as include environmental and social factors in the analysis. We did, however, have problems with sample size. A power analysis (Dunlop et al., 2012) found that the sample size, using only focal follow data, was insufficient to confidently detect a significant change in behaviour. However, combining focal data with *ad lib* data improved the power to 0.9. Testing the effect of social context remained problematic due to the large number of social contexts therefore future studies should focus on achieving a more robust sample size per social group using the focal follow methodology, or focusing on a small number of social group types. These experiments show that sound exposure generates a measurable behavioural response, but different exposure metrics should be considered, and this will be useful in future experiments aiming to test the hearing range of humpback whales as well as testing the function of many different types of social sounds.

ACKNOWLEGMENTS

The authors thank everyone involved in the Humpback Acoustic Research Collaboration (HARC), in particular John Noad, Nicoletta Biassoni, Ceri Morris, Melinda Rekdahl and the numerous volunteers who donated their time and energy to this project. We also thank David Paton for his invaluable field expertise and Simon Blomberg for his advice on statistical analysis. The authors would also like to thank Michael Bryden for taking the time to review earlier drafts as well as the anonymous reviewers for providing valuable input.

FUNDING

This work was supported by The US Office of Naval Research, the Australian Defence Science and Technology Organisation, the Australian Marine Mammal Centre and the Joint Industry Program E&P Sound and Marine Life.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

REFERENCES

Alexander, R. D. (1961). Aggressiveness, territoriality, and sexual behavior in field crickets (Orthoptera: Cryllidae). *Behaviour* **17**, 130-223.

Au, W. W. L., Pack, A. A., Lammers, M. O., Herman, L. M., Deakos, M. H. and Andrews, K. (2006). Acoustic properties of humpback whale songs. *Journal of the Acoustical Society of America* **120**, 1103-1110.

Baker, C. S., and Herman, L. M. (1984). Aggressive behavior between humpback whales (*Megaptera novaeangliae*) wintering in Hawaiian waters. *Canadian Journal of Zoology* **62**, 1922–1937.

Bates, D., Maechler. M., Bolker, B. and the R Core team. (2011). Linear mixed-effects models using S4 classes. *R: A Language and Environment for Statistical Computing*, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2009, ISBN 3-900051-07-0, http://www.R-project.org.

Brown, M.R. & Corkeron, P.J. (1995). Pod characteristics of migrating humpback whales (*Megaptera novaeangliae*) off the east Australian coast. *Behaviour* **132** (3-4), 163-179.

Cato, D.H. (1997). Ambient sea noise in Australian waters. *Proceedings of the Fifth International Congress on Sound and Vibration*, Adelaide, Casual Productions, Adelaide, vol. 5, p. 2813-2818.

Clapham,P.J. (2000). The humpback whale. Seasonal feeding and breeding in a baleen whale. In: *Cetacean societies: Field studies of dolphins and whales*, (ed. J. Mann, R. C. Conner, P. L. Tyack and H. Whitehead), pp 173-196, University of Chicago Press, Chicago

Darling, J. D. (1983). Migration, abundance and behavior of Hawaiian humpback whales, *Megaptera novaeangliae*. Ph.D Thesis, Santa Cruz: University of California. p.147.

Deecke, V. B. (2006). Studying marine mammal cognition in the wild: A review of four decades of playback experiments. *Aquatic Mammals* **32**, 461-482.

Dunlop, R. A., Cato, D. H. and Noad, M. J. (2012). Behavioral response studies: problems with statistical power. In *The Effects of Noise on Aquatic Life, Advances in Experimental Medicine and Biology Vol 730,* eds. A. N. Popper and A. Hawkins, pp. 293-297, Springer Science + Business Media, LLC 2012

Dunlop, R. A., Cato, D. H. and Noad, M. J. (2010). Your attention please: increasing ambient noise levels elicits a change in communication behaviour in humpback whales

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

(*Megaptera novaeangliae*). *Proceedings of the Royal Society B: Biological Sciences* **277**, 2521 – 2529.

Dunlop, R. A., Cato, D. H. and Noad, M. J. (2008). Non-song acoustic communication in migrating humpback whales (*Megaptera novaeangliae*). *Marine Mammal Science* **24**, 613-629.

Dunlop, R. A., Noad, M. J., Cato, D. H. and Stokes, D. (2007). The social vocalization repertoire of east Australian migrating humpback whales (*Megaptera novaeangliae*). *Journal of the Acoustical Society of America* **122**, 2893-2905.

Espmark, Y. (1971). Individual recognition by voice in reindeer mother-young relationship. Field observations and playback experiments. *Behaviour* **40**, 295-301.

Falls, J. B. (1992). Playback - a Historical-Perspective. *Playback and Studies of Animals Communication* **228**, 11-33.

Ficken, M. S. and Ficken, R. W. (1970). Responses of 4 Warbler Species to Playback of Their 2 Song Types. *Auk* **87**, 296-&.

Fish, J. F. (1968). Effect of Sound Playback on Toadfish (*Opsanus Tau*). *American Zoologist* **8**, 691-&.

Frankel, A. S. and Clark, C. W. (1998). Results of low-frequency playback of M-sequence noise to humpback whales, *Megaptera novaeangliae*, in Hawai'i. *Canadian Journal of Zoology* **76**, 521-535.

Frankel, A. S. and Clark, C. W. (2000). Behavioural responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals. *Journal of the Acoustical Society of America*, **108,**1930-1937.

Frankel, A. S., Clark, C. W., Herman, L. M. and Gabriele, C. M. (1995). Spatial distribution, habitat utilisation, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawai'i, determined using acoustic and visual techniques. *Canadian Journal of Zoology* **73**(6), 1134 – 1146.

Haskell, P. T. (1957). The Influence of Flight Noise on Behaviour in the Desert Locust Schistocerca-Gregaria (Forsk). *Journal of Insect Physiology* **1**, 52-&.

Johnson, M. P. and Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering* **28**, 3-12.

Mellinger, D. K. (2001). Ishmael 1.0 User's Guide." NOAA Technical Memorandum OAR PMEL-120, 2001.

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 airguns on the foraging behavior of sperm whales in the Gulf of Mexico. *Deep-Sea Research I* **56**, 1168-1181.

Mobley Jr., J. R., Herman, L. M. and Frankel, A. S. (1988). Responses of wintering humpback whales (*Megaptera novaeangliae*) to playback of recordings of winter and summer vocalisations and of synthetic sound. *Behavioral Ecology and Sociobiology* **23**, 211-223.

Morgan, D. W. (1970). Reactions of White Whale to Playback of Their Natural Sounds Underwater. *American Zoologist* **10**, 294.

Payne, K. and Payne, R. (1985). Large scale changes over 19 years in song of humpback whales in Bermuda. *Zeitschrift für Tierpsychologie* **68**, 89-114.

Payne, R. S. (1978). Behavior and vocalizations of humpback whales (*Megaptera* sp.). In: Report on a workshop on problems related to humpback whales (*Megaptera novaeangliae*) in Hawai'i, ed. K.S. Norris & R.R. Reeves. NTIS, PB- 280 794, Report No. MMC-77/03., pp. 56-78.

Payne, R. S. and McVay, S. (1971). Songs of humpback whales. *Science* **173**, 585-597.

Richardson, W. J., Greene Jr, C. R., Malme, C. I. and Thomson, D. H. (1995). *Marine Mammals and Noise*. San Diego, California, Academic Press.

Roche, J. C. (1966). Bird sound: First results of experimental research based on "playback". *Alauda Rev Int Ornithol* **34**, 55-62.

Smith, J. N., Goldizen, A. W., Dunlop, R. A. and Noad, M. J. (2008). Songs of male humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions. *Animal Behaviour* **76**, 467-477.

Smith, J. N., Grantham, H. S., Gales, N., Double, M. C., Noad, M. J. and Paton, D. (2012). Identification of humpback whale breeding and calving habitat in the Great Barrier Reef. *Marine Ecology Progress Series* **447**, 259-272.

Southall, B. L., Bowles, A. E. , Ellison, W. T., Finneran, J. J, Gentry, R. L., 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: Initial Scientific Recommendations. *Aquatic Mammals* **33**, 1-521.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Tyack, P. (1981). Interactions between singing Hawaiian humpback whales and conspecifics nearby. *Behavioural Ecology and Sociobiology* **8**, 105-116.

Tyack, P. L. (1983). Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds. *Behavioral Ecology and Sociobiology* **13**, 49-55.

Tyack, P. L. and Clark, C. W. (2000). Communication and acoustic behaviour of dolphins and whales. In *Hearing by Whales and Dolphins*, eds. W. W. L. Au A. N. Popper and R. R. Fay), pp. 156-224. New York: Springer-Verlag.

Tyack, P. L., & Whitehead, H. (1983). Male competition in large groups of humpback whales. *Behaviour* **83**, 132-154.

Verner, J. and Milligan, M. M. (1971). Responses of Male White-Crowned Sparrows to Playback of Recorded Songs. *Condor* **73**, 56-&.

Watkins, W. A. and Schevill, W. E. (1968). Underwater Playback of Their Own Sounds to Leptonychotes (Weddell Seals). *Journal of Mammalogy* **49**, 287-&.

Zoidis, A. M., Smultea, M. A., Frankel, A. S., Hopkins, J. L., Day, A., McFarland, A. S., Whitt, A. D. and Fertl, D. (2008). Vocalizations produced by humpback whale (*Megaptera novaeangliae*) calves recorded in Hawaii. *Journal of the Acoustical Society of America* **123**, 1737-1746.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

FIGURE LEGENDS

Fig. 1. Graph showing the dive profile (metres) (a) and direction (degrees from true north) (b) of a tagged mother from a female-calf group exposed to ‘social sounds’.

Fig. 2. Mean (± SEM) number of surface intervals per experimental period (n = 53) and dive time (in focally followed groups, n = 20) during the experiment for baseline groups and groups exposed to ‘social sounds’ and ‘tones’. Data points are slightly offset along the experimental period axis for clarity.

Fig. 3. Illustrating the course travelled (a), number of surfacing events (b) and during exposure of humpback whale groups to ‘tones’ as a function of the received level (triangles) and received signal-to-noise ratio (circles) at the start of the exposure period.

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

TABLE LEGENDS

Table 1. The sample size of different measured cohorts for the two visual platforms of observation

Table 2. Table illustrating the number of groups used in the BRS analysis for each platform of observation and each exposure.

Table 3. Estimated difference response measures (including standard error, t value, MCMC estimation of difference and p value) comparing the two sampling methodologies, *ad lib* and focal follow, on 15 groups sampled using both methodologies simultaneously (15 groups, 84 observations). Focal follow measures were used as the 'baseline' with which to compare the difference in *ad lib* measures.

Table 4. Comparison of the four different response models for course travelled and number of surface intervals (including model degrees of freedom, AIC, χ^2, χ^2 degrees of freedom and P values. NS designates that the model is not significantly improved from the previous model. Bold designates the 'best' model in terms of the lowest AIC score (lowest residual deviance).

	Focally followed	*Ad lib* sampling
Singletons	0	13 (includes 8 singers)
Female-calf	6	11 (1 tagged)
Female, calf and escort	6	11 (includes 1 singing escort)
Adult pairs	4	8 (including 2 singer/non-singer pairs)
Plus three adults	4	4 (adult trios and female-calf-escorts)

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Platform of observation	Sample size			
	None	Silence	Social sounds	Tones
DTAG	0	0	1	0
Focal follow	6	0	8	6
Ad lib sampling	19	5	11	12

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

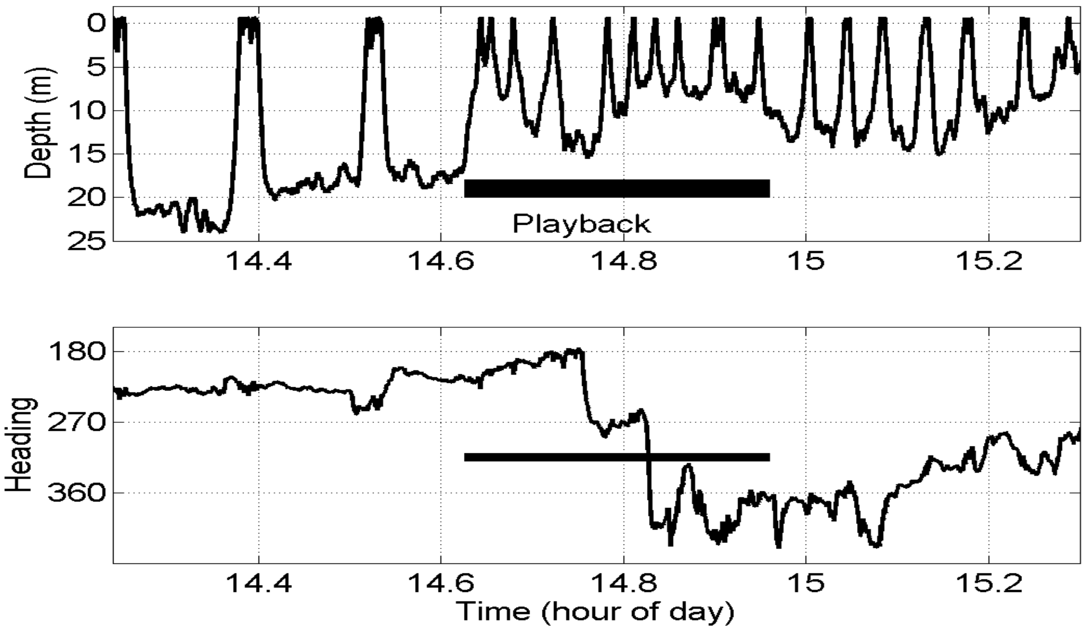
Response measure (per experimental period)		Estimate	SE	t value	MCMC mean	p-value
Course travelled (degrees)	Intercept (focal)	170.30	6.06			
	Ad lib	2.00	6.40	0.31	2.01	NS
Variation in course travelled (degrees)	Intercept (focal)	2.92	0.15			
	Ad lib	-0.07	0.16	-0.41	-0.07	NS
Speed (km/h)	Intercept (focal)	5.48	0.34			
	Ad lib	-0.29	0.28	-0.41	-0.28	NS
‘Long’ dive time (s)	Intercept (focal)	261.54	14.47			
	Ad lib	-31.35	15.48	-0.41	-35.34	0.03
Number of surfacing intervals	Intercept (focal)	3.33	0.27			
	Ad lib	-0.12	0.26	-0.46	-0.12	NS
Surface interval time (s)	Intercept (focal)	116.85	17.45			
	Ad lib	-51.82	19.63	-2.64	-50.58	0.01
No of ‘blow only’ surface intervals	Intercept (focal)	1.97	0.26			
	Ad lib	0.59	0.31	1.93	0.59	NS
No of ‘surface active’ surface intervals	Intercept (focal)	0.34	1.96			
	Ad lib	-0.71	0.25	-2.84	-0.71	0.005
No of observations	Intercept (focal)	11.33	1.46			
	Ad lib	-5.34	1.16	-4.33	-2.56	0.0001

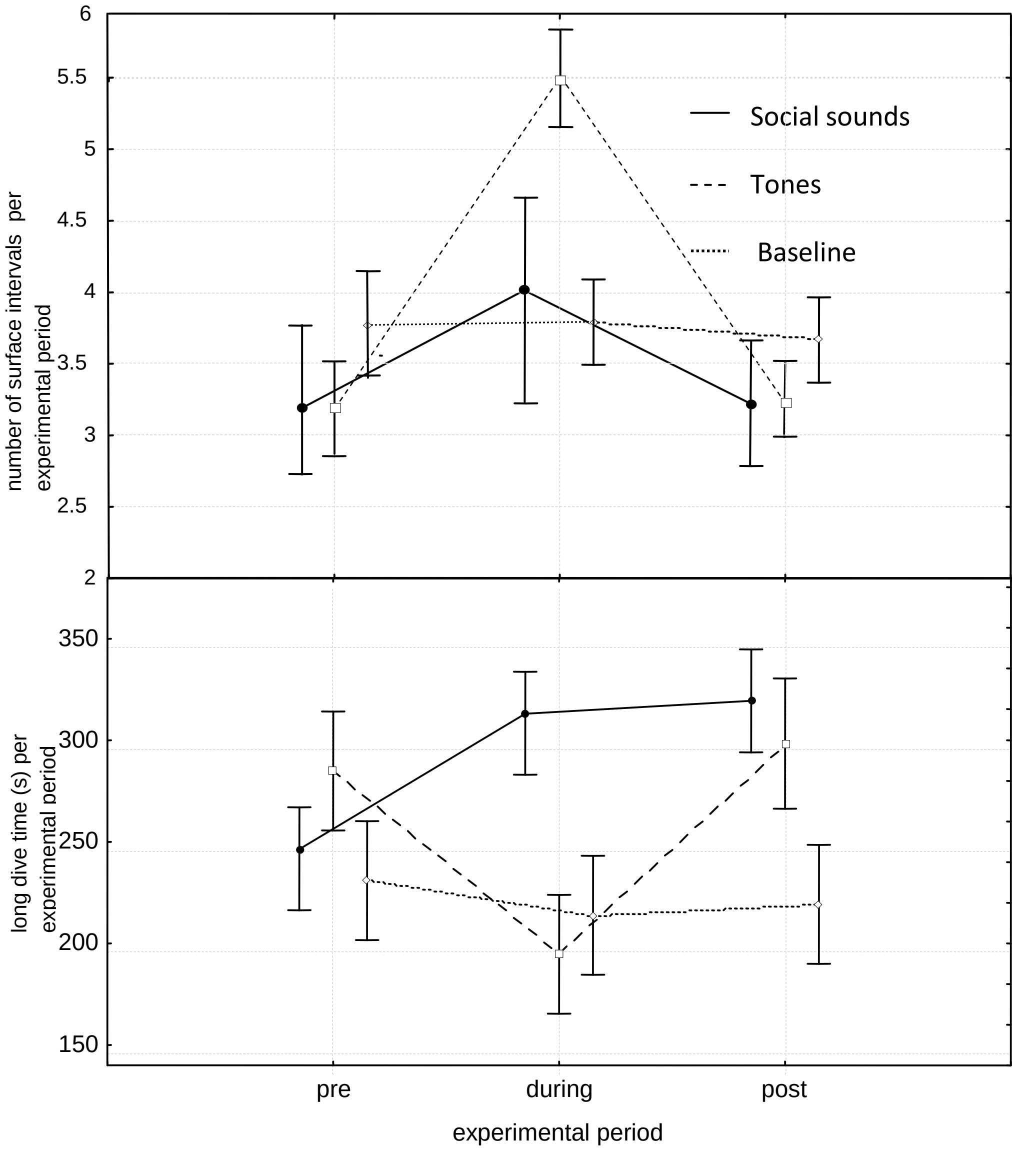
The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

Course Travelled (n = 37, full data)		*Df*	AIC	χ^2	χ^2_{df}	*P-value*
stimulus* experimental period		8	1167			
stimulus* experimental period* proximity	1 v 2	14	1166	12.5	6	=0.05
stimulus* experimental period* *RL*	2 v 3	14	1167	0	0	NS
stimulus* experimental period* *SNR*	**3 v 4**	**14**	**1163**	**4.2**	**0**	**<0.0001**
Course Travelled (n = 23, ‘audible’ data)						
stimulus* experimental period		8	716			
stimulus* experimental period* proximity	1 v 2	14	718	9.9	6	NS
stimulus* experimental period* *RL*	2 v 3	14	715	3.1	0	NS
stimulus* experimental period* *SNR*	3 v 4	14	723	0	0	NS
Number of surface intervals (n = 37, full data)						
stimulus* experimental period		8	-98			
stimulus* experimental period* proximity	1 v 2	14	-110	23.6	6	0.0006
stimulus* experimental period* *RL*	**2 v 3**	**14**	**-120**	**10.2**	**0**	**<0.0001**
stimulus* experimental period* *SNR*	**3 v 4**	**14**	**-120**	**0**	**0**	**NS**
Number of surface intervals (n = 23, ‘audible’ data)						
stimulus* experimental period		11	-65			
stimulus* experimental period* proximity	1 v 2	14	-61	8.3	3	NS
stimulus* experimental period* *RL*	2 v 3	14	-63	1.6	0	NS
stimulus* experimental period* *SNR*	**3 v 4**	**14**	**-78**	**15.9**	**0**	**<0.0001**
Long dive time (n = 14, focal follow data)						
stimulus* experimental period		11	633			
stimulus*experimental period*proximity	**1 v 2**	**14**	**435**	**203.6**	**3**	**<0.0001**
stimulus* experimental period* *RL*	2 v 3	14	440	0	0	NS
stimulus* experimental period* *SNR*	3 v 4	14	443	0	0	NS

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT





The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT

The Journal of Experimental Biology – ACCEPTED AUTHOR MANUSCRIPT