

## Individual, ecological, and anthropogenic influences on activity budgets of long-finned pilot whales

S. ISOJUNNO,<sup>1,2,†</sup> D. SADYKOVA,<sup>3,4</sup> S. DERUITER,<sup>5</sup> C. CURÉ,<sup>6</sup> F. VISSER,<sup>7,8</sup> L. THOMAS,<sup>2</sup>  
P. J. O. MILLER,<sup>1</sup> AND C. M. HARRIS<sup>2</sup>

<sup>1</sup>Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, KY16 8LB UK

<sup>2</sup>Centre for Research into Ecological & Environmental Modelling (CREEM), The Observatory, Buchanan Gardens, University of St Andrews, St Andrews, Fife KY16 9LZ UK

<sup>3</sup>Institute of Biological and Environmental Sciences, School of Biological Sciences, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ UK

<sup>4</sup>School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL UK

<sup>5</sup>Mathematics and Statistics Department, Calvin College, Grand Rapids, Michigan 49546 USA

<sup>6</sup>Cerema, DTer Est, Acoustics Group, F-67035 Strasbourg Cedex 2, France

<sup>7</sup>Kelp Marine Research, Loniussstraat 9, 1624 CJ Hoorn, The Netherlands

<sup>8</sup>Behavioural Biology, Institute of Biology, Leiden University, P.O. Box 9505, 2300 RA Leiden, The Netherlands

**Citation:** Isojunno, S., D. Sadykova, S. DeRuiter, C. Curé, F. Visser, L. Thomas, P. J. O. Miller, and C. M. Harris. 2017. Individual, ecological, and anthropogenic influences on activity budgets of long-finned pilot whales. *Ecosphere* 8(12): e02044. 10.1002/ecs2.2044

**Abstract.** Time allocation to different activities and habitats enables individuals to modulate their perceived risks and access to resources and can reveal important trade-offs between fitness-enhancing activities (e.g., feeding vs. social behavior). Species with long reproductive cycles and high parental investment, such as marine mammals, rely on such behavioral plasticity to cope with rapid environmental change, including anthropogenic stressors. We quantified activity budgets of free-ranging long-finned pilot whales in order to assess individual time trade-offs between foraging and other behaviors in different individual and ecological contexts, and during experimental sound exposures. The experiments included 1–2 and 6–7 kHz naval sonar exposures (a potential anthropogenic stressor), playback of killer whale (a potential predator/competitor) vocalizations, and negative controls. We combined multiple time series data from digital acoustic recording tags (DTAG) as well as group-level social behavior data from visual observations of tagged whales at the surface. The data were classified into near-surface behaviors and dive types (using a hidden Markov model for dive transitions) and aggregated into time budgets. On average, individuals (N = 19) spent most of their time (69%) resting and transiting near surface, 21% in shallow dives (depth <40 m), and only 10% of their time in deep foraging dives, of which 65% reached a depth 10 m from the sea bottom. Individuals in the largest of three body size classes or accompanied by calves tended to spend more time foraging than others. Simultaneous tagging of pairs of individuals showed that up to 50% of the activity budget was synchronized between conspecifics with decreased synchrony during foraging periods. Individuals spent less time foraging when forming larger non-vocal aggregations of individuals in late afternoons, and more time foraging when in the mid-range of water depths (300–400 m) available in the study area (50–700 m). Individuals reduced foraging time by 83% (29–96%) during their first exposure to sonar, but not during killer whale sound playbacks. A relative increase in foraging during repeat sonar exposures indicated habituation or change in response tactic. We discuss the possible adaptive value of these trade-offs in time allocation to reduce individual conflict while maintaining benefits of group living.

**Key words:** activity synchrony; anthropogenic noise; benthic habitat; deep-diving mammal; DTAG; ethogram; *Globicephala melas*; hidden Markov model; naval sonar.

**Received** 11 July 2017; revised 16 October 2017; accepted 23 October 2017. Corresponding Editor: Hunter Lenihan.

**Copyright:** © 2017 Isojunno et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** si66@st-andrews.ac.uk

## INTRODUCTION

Animals have evolved behavioral response and learning strategies to cope with both stable and variable aspects of their environment. For behavioral responses to be adaptive, individuals must assess the cost–benefit of behavioral change against perceived risk and opportunity in their individual (e.g., body condition, age), social (e.g., group size), and environmental (e.g., resource quality, location) contexts. This assessment is inherently uncertain, particularly in unfamiliar or unpredictable environments, and can therefore lead to either adaptive or maladaptive behavioral responses. Human-induced rapid environmental change, such as noise pollution, may further increase this uncertainty (Sih 2013) and, similar to predation risk (Frid and Dill 2002), may influence an individual's cost–benefit assessment and subsequent investment of time and energy in different behavioral options. If persistent, such behavioral decisions may have consequences for individual fitness and may ultimately impact population viability (Gill et al. 2001, Frid and Dill 2002, Beale 2007, Dunbar et al. 2009, New et al. 2014). A promising approach to assess biologically significant outcomes of behavior change is to quantify their cost as the level of time and energy trade-offs that individuals make in different risk–reward contexts (Houston et al. 2012, Isojunno et al. 2016). With the development of animal-borne data loggers, there has been increasing scope to measure such costs for free-ranging animals where individual behavior can be linked to realistic environmental contexts (e.g., prey availability; Friedlaender et al. 2016) and thus directly contributing to conservation science.

Time allocation to different activities and habitats is a key behavior tactic that individuals use to modulate their level of exposure to different types of resources and risks (Brown and Kotler 2004). For example, optimal foragers should give up foraging and switch to searching when the energetic, predation, and missed opportunity costs exceed the benefit of foraging in a food patch (Brown and Kotler 2004). Time allocation may be an especially important constraint in social species where individuals need to decide when to switch from social interactions to other crucial activities such as foraging or resting (Pollard and Blumstein 2008, Dunbar et al. 2009), and when to synchronize behavior

with conspecifics (Conradt and Roper 2005, Sueur et al. 2011). Synchronized behavior can incur “consensus” costs for individuals that have different optimal time budgets from their group members due, for instance, to differential energy requirements (Côte et al. 1997, Conradt and Roper 2005). Those types of social trade-offs emerge in social foraging behavior where individual foraging can be influenced by the cues or signs of conspecific foragers (Galef and Giraldeau 2001), incurring both benefits (e.g., increased ability to find food) and costs (e.g., competition) (Marshall et al. 2012). However, behavioral studies have focused on group-level rather than individual-level time budgets for social species (Marshall et al. 2012) with few studies focusing on social influences on the timing of individual foraging (Galef and Giraldeau 2001). Moreover, while studies have reported the influence of individual, social, or environmental contexts on daily time budgets over periods of months, variation within shorter timescales such as due to diurnal cycles is still relatively poorly understood (Marshall et al. 2012).

Deep-diving marine mammals have to balance the energetic benefit of foraging against the time, energetic, and physiological cost of diving to depth (Boyd 1997, Kooyman and Ponganis 1998). Species such as sperm whales (*Physeter macrocephalus*) that form cohesive social groups at surface in between foraging dives (Whitehead 1996, Gero et al. 2009) may have the added trade-off between individually optimized feeding opportunities and the need to maintain or regain social cohesion. The long-finned pilot whale (*Globicephala melas*) is found in both shelf-edge and deep-water habitats in temperate and sub-polar waters of the North Atlantic (Abend and Smith 1999) and the Southern Ocean (Van Waerebeck et al. 2010), while a closely related congener species, the short-finned pilot whale (*Globicephala macrorhynchus*), inhabits warm temperate and subtropical waters. Both species feed on squid (Desportes and Mouritsen 1988, Gannon et al. 1997, Mintzer et al. 2008), make deep foraging dives (>500 m; Baird et al. 2002, Aguilar de Soto et al. 2008), and live in social groups that are thought to consist of related females and males (Amos et al. 1993, Ottensmeyer and Whitehead 2003, de Stephanis et al. 2008, Alves et al. 2013b). The timing of the foraging periods, but not necessarily individual foraging dives, appears temporally synchronized (Visser et al. 2014); long-finned

pilot whales have also been shown to perform closely synchronized surface and underwater movements (Senigaglia and Whitehead 2012, Aoki et al. 2013). Similar to other toothed whales, their echolocation clicks, with terminal series of rapid clicks known as buzzes indicating prey capture attempts (Miller et al. 2004), can be used to indicate their foraging effort. Initial tagging studies with long-finned pilot whales in the Ligurian Sea indicated that foraging occurred during nighttime, presumably to match the movements of their vertically migrating prey (Baird et al. 2002). In short-finned pilot whales, daytime dives were deeper and more likely to contain high-speed sprinting behavior (Aguilar Soto et al. 2008).

Toothed whales such as pilot whales are thought to be especially vulnerable to anthropogenic noise pollution as they use sound both to search for food (echolocation signals) and to maintain social contact with conspecifics (Southall et al. 2007). Navy sonar is of a particular concern, due to links to cetacean mass mortality events (D'Amico et al. 2009) and behavioral changes that may have consequences to individual survival and fitness in several marine mammal species (Southall et al. 2016). Long-finned pilot whales have been reported to have a wide range of behavioral responses to naval sonar, including avoidance, cessation of foraging, and changes in vocal and social behavior (Rendell and Gordon 1999, Miller et al. 2012, Sivle et al. 2012, Antunes et al. 2014, Visser et al. 2016). In some cases, those responses were only observed at high received levels of sonar (Antunes et al. 2014), indicating that they may not be as sensitive to disturbance from sonar as other toothed whales. Long-finned pilot whales have also been shown to be attracted to playbacks of killer whale (a potential predator/food competitor) sounds in an apparent mobbing response (Curé et al. 2012). However, the costs of such responses have not been quantified in terms of individual-level time trade-offs.

Our overall objective was to assess in free-ranging long-finned pilot whales the individual investment in time spent foraging, given other behavioral options, in diverse contexts that we expected to influence the perceived cost-benefit and risk of foraging. We quantified an ethogram of tagged whales in order to indicate how foraging time may be traded off for other functional behaviors. Specifically, we expected individual

time spent foraging to vary with the following internal and external factors: (1) individual body size and association with a calf, due to increased energetic requirements; (2) time of day and water depth, due to availability and accessibility of prey; (3) social context, due to a need to maintain social contact but reduce inter-specific competition for food; and (4) experimental sound exposures, due to a trade-off between foraging and safety; that is, the perceived cost or risk from the signals exceeds the benefit of foraging in a given context. The experimental exposures included 1- to 2- and 6- to 7-kHz naval sonar (an anthropogenic and potentially novel signal) and playback of killer whale vocalizations, as well as negative controls for both an approaching and a nearby stationary source.

## METHODS

### Overview

The first part of the analysis aimed to construct a complete ethogram for long-finned pilot whales and quantify transition probabilities between behavior states to indicate how foraging behavior may be traded off for other functional behaviors, such as traveling, resting, or socializing. The second part of the analysis aimed to quantify any variation in foraging time allocation in relation to different intrinsic and extrinsic factors. An overview of the stages of the analysis is given in Appendix S1: Fig. S2.

The ethogram was constructed by combining data from 19 tag records (including data on depth, acceleration, magnetic field, and acoustics) with surface visual observations of the social context of tagged whales. The multivariate data were summarized between individual breath times (inter-breath interval, IBI) detected from the tag, providing natural break points for potential changes in behavior state. Dive depth and duration thresholds were used to define a subset of the IBIs as dives, which were the main focus of the analysis. Dive thresholds are typically defined a priori depending upon desired analysis resolution and/or available resolution from onboard pressure sensors (e.g., dives defined as longer duration than 6 s in Mate et al. [2005], or deeper than 5 m in Aoki et al. [2013]). To inform these thresholds by behavior, we used mixture models to classify IBI data variables that we expected to reflect the

animals' need to take repeated breaths. The shallower IBI class estimated by the mixture model was used to characterize the IBI and maximum depth distribution of near-surface movements (NSMs) and select dive thresholds. The resulting NSMs and dives (separated by the thresholds) were classified further in a hidden Markov model (HMM) framework to account for the time series nature of the tag data and to quantify transition probabilities between behavior states. Separate HMMs were fit to dives and NSMs, because (1) there were many more NSMs than dives so a joint model would have been dominated by NSM sub-categories when our main focus was on identifying foraging events within dives, and (2) the most insightful data to classify within each category was different. The HMM for dives included rich multivariate time series in order to identify foraging dives and test the number of other distinct dive types. For NSMs, we were particularly interested in identifying horizontal travel because it may indicate changes in habitat preference and avoidance responses under disturbance. Near-surface movements were therefore further classified in a two-state HMM that included movement variables to identify any traveling behavior while the animals were near surface.

The HMMs for dives were specified with candidate covariates and random effects to allow for individual and temporal variation in transition probabilities between behavior states. However, variation in the dive-by-dive transition probabilities does not necessarily lead to variation in total time allocation that also includes near-surface behavior. Therefore, the classified near-surface and dive behavior states were aggregated into time budgets, and the effect on time spent foraging of sound exposures (e.g., naval sonar) and other explanatory variables representing different intrinsic (e.g., body size) and extrinsic factors (e.g., water depth) was tested via binomial regression in a second part of the analysis (Appendix S1: Fig. S2). The effect of social context was addressed by analyzing series of activity states from pairs of whales tagged in close vicinity of each other. We tested our expectation that shallow dives would be more synchronous than deep foraging dives due to lower consensus costs, that is, reduced conflict between conspecifics when outside a food competition context. We provide details of each step in the sections below.

### Data collection

Data were collected from 19 long-finned pilot whales, of which 18 were tagged with audio- and movement-recording data loggers using suction cups (DTAG; Johnson et al. 2009) and one whale (gm10\_144a) was tagged with a logger that recorded three-axis acceleration, depth, and speed but neither magnetic field nor acoustics (PD3GT Little Leonardo, Aoki et al. 2013). Magnetometer data were also missing for one tag record in 2008 (gm08\_154d) due to sensor failure. On four occasions, a second whale was tagged, allowing us to investigate synchrony in behavior between two animals tagged in the same area. The whales were tagged in the Vestfjord basin off Lofoten in northern Norway (66–70° N latitude) during the spring and summer 2008–2014. The field protocol included (1) tagging the whale from a small rigid-hulled inflatable boat (RHIB) using a handheld pole and a suction-cup attachment, (2) visual and VHF (Very high frequency) tracking of the tagged whale, and (3) recovery of the released tag (after 10–15 h of recording).

The location and social context of the tagged whale were sampled every two minutes according to a standardized visual observation protocol in 2009–2014 (Miller et al. 2011, Visser et al. 2014). If more than one animal was tagged at a time, the first one with a successful tag attachment high on the body (i.e., with a good VHF signal) was selected as the focal whale for visual tracking and surface behavior data collection. The position of the focal whale was estimated using field-estimated range and bearing relative to the course and GPS position of the observation vessel. The observation vessel aimed to maintain 100–400 m range to the focal whale. Visual data were collected at the level of the focal “group,” defined as individuals in closer proximity to the tagged individual and each other than other individuals in the area. Group size, spacing (distance between individuals, measured in body lengths), presence/absence of calves, and distance to the nearest other group were recorded for the focal group. A full definition for these behavioral variables is provided in Visser et al. (2014).

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.

### *Experimental exposure procedures*

The exposure experiments were designed and conducted within the 3S (Sea mammals, Sonar, Safety) research project. The full experimental protocol is described in Miller et al. (2011, 2012) and in Curé et al. (2012) and only briefly summarized here.

Tagged whales were exposed to blocks of transmission (exposure sessions) of two or three of the following types of sonar: (1) mid-frequency active sonar (MFAS) 6- to 7-kHz hyperbolic upsweep, (2) low-frequency active sonar (LFAS) 1- to 2-kHz hyperbolic upsweep, or (3) LFAS 1- to 2-kHz hyperbolic downsweep. Sonar signals were 1 s in duration and were transmitted at 20-s intervals. Each sonar exposure session lasted 25-80 min and included only one sonar signal type, with source levels increasing over the first 10 min of the exposure session ("ramp-up" protocol). The towed source (SOCRATES, TNO, The Netherlands) was towed toward the whale subject at a depth of about 55 m (range 35–100 m), and source levels (dB re 1  $\mu$ Pa m) ranged from 152 to 214 dB for LFAS and from 158 to 199 dB for MFAS. Turns toward the tagged whale were ceased once the source vessel was within 1 km of the tagged whale. The sonar source was towed but not transmitting during no-sonar control approaches in order to separate potential effects of the approaching source from effects of sonar. The source ship was the 55 m R/V H.U. Sverdrup II. The order of signal type was changed across tag deployments to enable evaluation of order effects, and all exposure and no-sonar control sessions had at least an hour between them. The received level of the sonar signals was estimated as the maximum sound pressure level over a 200-ms window ( $SPL_{max}$ ; dB re 1  $\mu$ Pa; Miller et al. 2011).

In addition, sound playback experiments were conducted from a small motor boat (<10 m) that was stationed at ~800 m range from the tagged whale at the start of each playback and was allowed to drift over the course of the playback (Curé et al. 2012). The stimuli included LFAS sounds that represented a less powerful and stationary LFAS exposure to be contrasted with the

towed LFAS sonar approaches, and natural sequences of killer whale sounds recordings designed to simulate a natural high-level disturbance (predator/competition) context providing a positive control to the sonar exposures. Further, playbacks of broadband noise were conducted as a negative control for the playback stimuli (Curé et al. 2012). The broadband noise control playbacks were prepared from the non-calling periods of the killer whale DTAG recordings; they included ambient noise and flow noise from the tag, amplified to get an average root-mean-square power equal to the killer whale calls. These playback stimuli were 15 min in duration and were broadcast at source levels of 145–151 dB re 1  $\mu$ Pa m.

For analysis, data were excluded from the beginning of the tag record until the end of tagging operations (when the boat used for pole-tagging was no longer active in the vicinity of the whale; Isojunno and Miller 2015). The data after tagging, but preceding any experimental control or sound exposure, were considered to be baseline data. To minimize potential research effects related to vessel proximity, the observation vessel consistently aimed to keep a distance of 200 m of the tagged whale, avoiding any close (<100 m) and direct approaches. Our analysis therefore assumes that any effect of the observation vessel was negligible and constant across the baseline and exposure periods.

### *Behavioral data*

Depth, pitch, and roll data (derived following Johnson and Tyack 2003, decimated at 5 Hz) were assessed visually in a custom-built program in MATLAB 8.6 (MathWorks, Natick, Massachusetts, USA) to mark breath times in the time series. Breaths were characterized by an arch in the pitch signature near the sea surface. Breath time was defined at the point at which the pitch was briefly horizontal. Visits near the sea surface with roll other than zero were not marked as breaths, and instead were considered part of the breath-holding interval between breaths. Near-surface behaviors with uncertainty about the number or timing of breaths were marked as "surface intervals." These intervals could include logging behaviors where animals were near-stationary at the surface while breathing. Movement and acoustic data extracted from the DTAG and concurrent visual observations were then summarized for each IBI.

Movement data calculated included dive duration (min), maximum depth (m), total vertical displacement (m), fluke stroke rate ( $\text{min}^{-1}$ ), circular variance of roll, circular variance of absolute value of pitch (to represent the vertical sinuosity of the dive; R package CircStats version 0.2-4 [Jammalamadaka and Sengupta 2001] within R [R Development Core Team 2017]), turning angle, and horizontal speed. Horizontal movement was summarized for each interval as the turning angle between tag-recorded headings at consecutive breath times (over a 0.5-s averaging window), and mean horizontal speed of the focal group between visually recorded locations. Fluke stroke rate was calculated using an automated detector based upon cyclic variation in pitch (Tyack et al. 2006), with detection parameters determined manually for each tag record by inspecting the magnitude of the stroke signals within the pitch record.

Acoustic data recorded by the DTAG at 192 kHz sampling rate, 16-bit resolution, were examined visually using spectrograms (4096 point FFT [Fast Fourier transform], 50% overlap) and aurally to record start and end times of clicking, buzzing, and social sounds. Each sound level was scored as “quiet” (1), “average” (2), or “loud” (3) relative to the perceived average amplitude of the tag record. Level 2 and 3 sounds were used in the analysis and interpreted as the vocal behavior of the tagged whale and whales closest to the tagged whale. These were summarized as the number of each type of sound, and the proportion of time that echolocation clicks, buzzes, or social sounds were recorded within each IBI (excluding overlapping time). In total, 40% of the acoustic data were not analyzed or not available, including the entire accelerometer tag (gm10\_144a).

Visual data were linked to each dive by summarizing the data within the two minutes following the whale returning to surface. Speed was calculated between each pair of visual track locations that were <5 min apart in time. Each three-minute interval was allocated a mean group speed, mean group size, and presence/absence of tight or very tight individual spacing within the focal group (i.e., <3 body lengths of one another).

For the four instances when two whales were tagged simultaneously and therefore there were time-overlapping DTAG records, an index was calculated representing the temporal synchrony of IBIs. The index was calculated as the overlapping

dive time divided by the combined duration of each pair (overlapping + non-overlapping dive time). Each focal dive was then associated with the non-focal dive with the most time overlap.

#### *Data on individual and environmental contexts*

Data on the context of individual tagged whales included a body size class and whether or not the tagged whale was associated with a calf. Association with a calf was recorded during field observations when an adult-sized animal was tightly paired (<1–3 body lengths) with a clearly smaller animal during the majority of its time at surface. To maximize our chances to record mother–calf associations rather than more short-term adult–calf associations, only associations that lasted for the entire deployment (>4 h) were included in the analysis. Body size class was determined by combining field estimates (small/medium/large adult), and where available, estimates of dorsal fin size from good quality photographs of the tag attached to the dorsal fin of the whale. The base of the dorsal fin (Augusto et al. 2013) was measured in perpendicular photographs and scaled to known length of the tag.

Environmental data included water depth and slope, solar elevation, and time since solar noon. Bathymetry data were obtained from the high-resolution Marine Primary Data (MPD) of the Norwegian Hydrographic Service. In order to associate one bathymetry data point with each dive (in 10-m contours), sighting coordinates were interpolated linearly at the midpoint of each dive. Missing values were assigned when dive data occurred outside the visual track. This matching was conducted using Universal Polar Stereographic projection in geographic information system software Manifold 8.0 (Manifold Software Limited, Central, Hong Kong). Slope was calculated by rasterizing the bathymetric contour map at  $800 \times 800$  m resolution, and then calculating the slope from the nearest neighbor pixels ( $3 \times 3$  window in Slope transform function, Manifold). Solar elevation (deg) and time since solar noon (h) were calculated for each IBI with respect to the dive start time using R package Maptools (algorithms provided by NOAA; Meeus 1991).

#### *Classification of near-surface movements*

The first step of the analysis aimed to select a dive depth and duration thresholds to separate

dives from stereotyped NSMs associated with the animals' need to take repeated breaths at the sea surface. Such surface movements were expected to be relatively stereotyped, with a short IBI, small vertical displacement, and a relatively constant position of the blow hole relative to the surface. These behaviors were characterized by fitting a two-state multivariate mixture model to the data (IBI, vertical displacement, and circular variance of roll). Vertical displacement, instead of maximum depth, was used to include information about both depth and the vertical sinuosity of the IBI. The model aimed to estimate two latent states: "NSMs" during surfacing, and dives. Each of the three data variables was assumed to have a mixture distribution where the parameters of the component distributions were dependent upon the latent state. Thus, by maximizing the joint likelihood of the three mixture distributions, we could estimate both the state-dependent parameters of each distribution (e.g., mean and variance of IBI during a dive) and the most likely state membership of each observation (i.e., surfacing vs. diving). Interbreath interval was fitted by a Weibull distribution (which allows for positive real values and right- or left-skewed distributions of IBIs during breathing vs. diving), vertical displacement, an exponential distribution, and circular variance of roll a beta distribution (which allows for values between 0 and 1 and is extremely flexible in terms of the shape of the distribution). The classified NSMs were used to define a dive depth and duration threshold for dives that were included in the HMM analysis. The selected thresholds were used (instead of the most likely states under the mixture model estimates) in order to remove state uncertainty in tag records that did not contain data on roll, and to provide a clear-cut definition of dives for future studies. The thresholds were selected using a percentile value for dive depth and duration within the NSMs classified by the mixture model; IBIs exceeding either the specified dive depth or duration were classified as dives; all others were NSMs. The percentile was selected to minimize the inclusion of NSMs in the subset of IBIs that would be considered as dives, and so we aimed primarily to minimize false positive rate (<0.05) while maximizing sensitivity of the thresholds to detect mixture model classified dives (Appendix S1: Fig. S3d, e).

The NSMs (those IBIs not exceeding the dive depth and dive duration thresholds) were further classified into near-surface traveling and non-traveling states in a two-state multivariate HMM (Zucchini et al. 2016). Since dives were removed from the time series to fit the model, state transition probabilities were assumed to be the same between NSMs occurring sequentially and NSMs that occurred immediately prior to and following a dive. The HMM included horizontal speed as well as turning angle, vertical speed, and fluke stroke rate to complement the sparse visual observation data. The three positive real-value data variables (horizontal speed, vertical speed, and fluke stroke rate) were assumed to follow a gamma distribution. Turning angles, which can have values between  $-180^\circ$  and  $180^\circ$ , were specified to have a von Mises distribution (circular normal distribution). For details about fitting mixture models and HMMs, please see the following section.

#### *Classification of dive types and analysis of dive transitions*

Dive types were classified by fitting multivariate HMMs (Zucchini et al. 2016) to the movement, acoustic, and visual data summarized for each dive. The dive summary metrics were modeled as state-dependent processes, and the probability of transition from one latent state (dive type) to the next state was described by a transition probability matrix (TPM). The dive summary metrics were selected to reflect the animals' diving effort (dive depth, duration, pitch variance), horizontal swimming effort (horizontal speed, turning angle), foraging behavior (presence/absence of echolocation), and social behavior (group size, presence/absence of social sounds, and presence/absence of tight spacing within the group). The presence/absence of pre-dive and post-dive surfacing (NSM) was also included in the model to provide information about transitioning between near-surface behavior and dives.

Similar to the mixture model, a parametric family of distributions was specified for each dive summary metric (Appendix S2). Covariates were included in the model to allow for variation in the Markov transition probabilities via multinomial logistic regression (Langrock et al. 2014, Zucchini et al. 2016). We also fitted models with a discrete random effect of individual whale, where the

transition probabilities of each individual were assumed to derive from one of  $K$  possible TPMs (Zucchini et al. 2016, DeRuiter et al. 2017). The number of states, number of discrete random effects, and covariates were selected based on information criteria (Appendix S1: Table S1). BIC (Bayesian information criterion) was used as the primary model selection criterion to avoid selection of overly complex models, but we also computed Akaike's information criterion (AIC) to assess sensitivity to the choice of the criterion. The likelihood of the HMM was computed following Zucchini et al. (2016) and DeRuiter et al. (2017). Please see Appendix S2 and supplementary R scripts in Data S1 for more details on the HMM structure and likelihood.

The negative log-likelihood of both mixture models and HMMs was minimized using the `nlm` function in R (package `stats`). Mixture distributions are multi-modal, and therefore, the minimization is sensitive to the choice of starting values. To check for multiple minima and to ensure the algorithm did not terminate at a local minimum, each model was fitted 50 times with different initial values and the stability of the resulting likelihoods was monitored visually. Initial values for the distributional parameters were calculated from random 10% subsets of the input data, based upon a mean for one-parameter distributions, and both mean and variance for two-parameter distributions. For the TPM of the HMM, covariate coefficients and the mixture weights were generated by sampling a uniform distribution.

For HMMs without random effects, we used a dynamic programming algorithm (Viterbi algorithm) to compute the most likely sequence of underlying hidden states given the parameters and the observations (Zucchini et al. 2016). For mixture models and HMMs with random effects, states were decoded by computing the likelihood of the multivariate observations during each IBI (NSM or dive) given the parameter estimates for each state, and assigning the IBI to the most likely state.

To avert the possibility of pseudo-replication (similar behavior by simultaneously tagged whales) affecting the model results, we fitted the model using data from one of each pair of simultaneously tagged whales for which visual observations were collected (excluding data from non-focal whales gm09\_137c, gm09\_138b, gm13\_169b, and

gm14\_180b). The fitted model estimates were used to predict the dive states in these tag records.

#### *Proportion of time spent foraging*

The explanatory variables for the probability of transition from one dive type to the next in the HMM did not test changes in time budgets, as it did not account for potential changes in average time spent in different dives or at surface. We therefore also modeled the time spent in a dive type as a proportion of total time in consecutive time bins. The proportion of time was modeled as a binomial response variable in a generalized additive mixed model (GAMM; package `mgcv` in R; Wood 2004). The response variable was the proportion of time spent in the dive type that was most indicative of foraging (deep dive depth, presence of echolocation clicks). To account for any serial correlation in the time spent foraging, the proportion of time spent foraging in previous time bin (PRE.Foraging) was included as a covariate. Each experimental exposure was included as a single time bin in the data. Fifteen minutes was added to the killer whale sound playback periods, as previous analyses of these data have shown behavioral changes to last at least 15 min into the post-exposure period (Visser et al. 2016). Baseline and post-exposure periods, which could last for several hours, were binned into shorter time intervals. The duration of the baseline and post-exposure time bin was selected by applying the univariable model with PRE.Foraging to data with an increasing bin length. The duration was selected as the shortest time bin that removed any serial correlation in model residuals. To account for the variable bin lengths, the number of binomial trials in the model was specified as the duration of the time window divided by the mean foraging dive duration in the baseline (i.e., if the time bin encompassed an average foraging dive exactly, this would be represented in the model as 1 success over 1 trial). Two post-exposure time bins following each exposure were removed from the analysis.

Candidate covariates included mean water depth (m), time since solar noon (h), and presence/absence of different types of exposures (Table 1). To allow for non-linear relationships with the response variable, water depth and time since solar noon were specified as univariate smooths (cubic regression splines with shrinkage



Table 1. Candidate covariates for HMM transition probabilities and proportion of time in binomial GAMMs.

Covariate	Model	Description
solarelev	HMM	Solar elevation (degrees)
solarnoon	GAMM	Time since solar noon (h)
water.depth	GAMM	Mean water depth (m)
PRE.Foraging	GAMM	Proportion of time spent foraging in previous time bin
ind.calf	HMM, GAMM	1 for individuals associated with a calf, 0 otherwise
ind.large	HMM, GAMM	1 for individual in the largest size category, 0 otherwise
ind.small	HMM, GAMM	1 for individual in the smallest size category, 0 otherwise
ind.LoC	HMM	1 for individuals with large body size or a calf, 0 otherwise
SIL	GAMM	1 for no-sonar approach, 0 otherwise
MFAS	GAMM	1 for MFAS approach, 0 otherwise
LFAS	HMM, GAMM	1 for LFAS approach, 0 otherwise
MLFAS	HMM, GAMM	1 for MFAS and LFAS sonar approach, 0 otherwise
RL_max	HMM	Maximum received level (SPL <sub>max</sub> ) of MFAS or LFAS during a dive
RL_145	GAMM	Proportion of time bin MLFAS signals received >145 dB (SPL <sub>max</sub> ; dB re 1 µPa)
prev.MLFAS	HMM, GAMM	1 for sonar and control approaches presented after a sonar approach, 0 otherwise
PB_BBN	GAMM	1 during broadband noise control playback, 0 otherwise
PB_KW_15	HMM, GAMM	1 during and 15 min after exposure of killer whale sound playbacks, 0 otherwise
PB_SON	GAMM	1 during LFAS sound playbacks, 0 otherwise
PBS	HMM	1 during PB_KW_15 or other playbacks of sounds, 0 otherwise including PB_BBN
SON	HMM, GAMM	1 during sonar approaches and LFAS sound playbacks, 0 otherwise
prev.PBS	GAMM	1 during playbacks presented after a playback of sounds, 0 otherwise
exposed	HMM	1 during any exposure and post-exposure, 0 during baseline
exposure	HMM	1 during any exposure, 0 during baseline and any post-exposure
exposureS	HMM	1 during any sound exposure, 0 during baseline and post-exposure
CTRL	HMM	1 during no-sonar approach and control playback, 0 otherwise

*Notes:* GAMM, generalized additive mixed model; HMM, hidden Markov model; LFAS, low-frequency active sonar; SPL, sound pressure levels. Not all experimental exposure types could be included separately in the HMM due to the large number of parameters in the model. We therefore combined the experimental sessions to several binary covariates. Water depth was not included as a candidate covariate in the HMM due to sparse horizontal tracking data.

penalty; Wood 2004). The maximum smooth basis dimension was set to 8 for time since solar noon and 5 for water depth. To account for variation in received acoustic level during the LFAS/MFAS approaches, we also included the proportion of time that sonar signals were received above 145 dB (SPL<sub>max</sub>; dB re 1 µPa), identified as the threshold for expert-identified behavioral responses (Miller et al. 2012, Harris et al. 2015). Order effects were tested by including two presence/absence covariates, one for towed approaches (prev.MLFAS) and the other for playbacks (prev.PBS), which were set to 1 (present) for control and sound exposures presented after a sonar approach or a playback of KW/sonar sounds, and zero (absent) otherwise. The focal individual was set as a random effect.

We fitted all combinations of covariates (including interactions between the order effects and exposure covariates) and selected the best model based upon its UBRE score (un-biased risk estimator; Wood 2004).

## RESULTS

### Data

In total, 19 tag records were analyzed; 15 of 19 tagged whales were exposed to naval sonar and/or control sound playbacks. A total of 153.9 h of tag data were analyzed, of which 70.1 h were baseline data (Table 2). All data were collected in waters up to 700 m deep, in Vestfjorden (Appendix S1: Fig. S1). Four pairs of animals were tagged close in time, and most of the time these paired tag records overlapped (Table 2; pair A: 6.9 h; pair B: 9.9 h; pair C: 6.6 h; and pair D: 8.0 h).

Across the 16 animals for which photographs were available, there was a reasonable concordance between the field-estimated body size class (small, medium, and large) and the size of the dorsal fin. Two field-estimated “medium” animals were re-classified to “small” (gm08\_150c, gm10\_152b), and one “large” animal was re-classified as “medium” (gm14\_180b). In the resulting classification, the base of the dorsal fin was estimated to be

Table 2. Summary of analyzed data.

ID (pair ID)	Body size	Calf	Experimental exposures	Total (h)	Baseline (h)
gm08_150c	Small	1	MFAS, LFAS	4.8	0.9
gm08_154d	Medium	1	LFAS, MFAS	6.5	2.0
gm08_159a	Large	0	NS CTRL, LFAS, MFAS, PB KW, PB KW	10.3	2.4
gm09_137b (A)	Medium	1		6.9	6.9
gm09_137c (A)	Small	0		7.6	7.6
gm09_138a (B)	Medium	0	LFAS, MFAS, NS CTRL, LFAS	9.9	3.2
gm09_138b (B)	Small	1	LFAS, MFAS, NS CTRL, LFAS, PB KW, PB KW	17.2	3.2
gm09_156b	Large	0	NS CTRL, LFAS, MFAS, LFAS, PB KW, PB KW	15.2	5.1
gm10_143a	Medium	1		8.8	8.8
gm10_144a	Large	0	PB CTRL, PB KW, PB CTRL, PB KW	5.4	1.6
gm10_152b	Small	0		1.6	1.6
gm10_157b	M-Large	1	PB CTRL, PB CTRL	11.1	10.5
gm10_158d	M-Large	0	PB CTRL, PB KW, PB CTRL, PB KW	7.4	2.9
gm13_137a	Small	0	PB HW, PB KW	6.6	2.8
gm13_149a	Large	0	PB LFAS, PB KW	5.0	2.0
gm13_169a (C)	Large	1	PB KW, PB LFAS, PB CTRL	6.6	2.1
gm13_169b (C)	Medium	0	PB KW, PB LFAS, PB CTRL	6.7	2.1
gm14_180a (D)	Large	0	PB LFAS, PB KW	8.0	2.4
gm14_180b (D)	M-Large	0	PB LFAS, PB KW	8.3	2.4
Sum				153.9	70.1

Notes: NS CTRL, no-sonar control approach; MFAS/LFAS, medium-/low-frequency (6–7 vs. 1–2 kHz) sonar approach; PB LFAS, near-stationary playback of sonar sounds; PB CTRL, broadband noise control playback; PB KW, playback of killer whale sounds. All tags were DTAGs, except for gm10\_144a which was an accelerometer tag. The time-overlapping pairs of dive records are labeled with A–D. As part of an objective for another research project, one trial playback of humpback whale sounds was conducted (PB HW). We included these data in the estimation of time budgets to maximize data sample (hidden Markov models), but the data were excluded from the binomial generalized additive mixed models that aimed primarily to quantify responses to sonar and killer whale sound exposures.

50–60 cm in small animals, 70–90 cm in medium animals, and 90–130 cm in large animals.

#### Classification of near-surface movements

The average duration and depth of the NSMs ( $n = 14024$ ) classified by the mixture model was 18.4 s ( $\pm 8.2$ ) and 2.7 m ( $\pm 1.1$ ; all values are mean  $\pm$  standard deviation). These IBIs had a low circular variation in roll ( $0.007 \pm 0.008$ ) compared to dives ( $0.158 \pm 0.204$ ). In total, 98% of NSMs were  $<37.8$  s long and 5.3 m deep. The 98% percentile had the maximum sensitivity (0.65) to detect dives with a false positive rate below 5% (0.038; Appendix S1: Fig. S3). Dives either longer or deeper than these thresholds are hereafter considered dives rather than surface behavior.

Almost a half (47%) of all NSMs ( $<37.8$  s and  $<5.3$  m) were classified to be near-surface traveling. The mean horizontal speed was elevated ( $1.9 \pm 0.79$  m/s) and the turning angle was smaller and less variable ( $9.40^\circ \pm 10.0^\circ$ ) during near-surface traveling compared to the other 53% of

NSMs ( $1.11 \pm 0.52$  m/s and  $21.2^\circ \pm 28.9^\circ$ ). Near-surface traveling also had a higher vertical speed and fluke stroke rate (Table 3; Appendix S1: Fig. S4). The model estimated 46% prevalence of traveling in the 40% of the surfacing data that did not have horizontal speed data.

#### Classification of dives and analysis of dive transitions

Hidden Markov model selection supported a time-homogeneous model (excluding covariates and random effects) with up to four different dive types (Appendix S1: Table S1). Including up to two or three discrete random-effects groups in the four-state HMM slightly decreased AIC (3 units) but significantly increased BIC (72 units). A reduction of 2 units is usually considered to be significant change in AIC, but combined with the large increase in BIC, these results indicate a relatively weak support for any random effects. No covariate further decreased either AIC or BIC of the four-state time-homogeneous HMM, except when individuals with a large body size and those

Table 3. Mean and standard deviation of data within each dive type and near-surface movement (NSM) state.

Variable	Foraging	Exploratory	Crowded	Directed	NSM travel	NSM other
Dive duration (min)	7.07 (2.91)	1.14 (0.64)	1.33 (0.89)	0.74 (0.18)	0.27 (0.12)	0.33 (0.13)
Dive depth (m)	301.6 (177.7)	9.9 (5)	10.6 (7)	5.3 (1.9)	2.7 (1)	2.6 (1)
Vertical speed (m/s)	1.4 (0.55)	0.4 (0.14)	0.38 (0.14)	0.33 (0.09)	0.41 (0.1)	0.32 (0.06)
Pitch variance	0.71 (0.2)	0.34 (0.24)	0.28 (0.23)	0.15 (0.13)	0.17 (0.1)	0.17 (0.1)
Roll variance	0.59 (0.27)	0.19 (0.23)	0.12 (0.19)	0.06 (0.13)	0.02 (0.06)	0.02 (0.06)
Horizontal speed (m/s)	1.16 (0.51)	1.11 (0.63)	1.3 (0.69)	1.53 (0.78)	1.9 (0.79)	1.11 (0.52)
Turning angle (°)	16.5 (16.5)	26.4 (34.6)	27.9 (35.5)	9.8 (9.2)	9.4 (10)	21.2 (28.9)
Stroke rate (s <sup>-1</sup> )	0.32 (0.11)	0.15 (0.12)	0.17 (0.12)	0.19 (0.14)	0.35 (0.15)	0.16 (0.1)
Group size	8.8 (6.1)	7.1 (3.4)	20.6 (8.7)	10.8 (3.5)	11 (6.2)	11.6 (8.7)
Tight group (0/1)	0.63 (0.48)	0.87 (0.34)	0.72 (0.45)	0.76 (0.43)	0.66 (0.48)	0.82 (0.38)
Social sounds (0/1)	0.98 (0.13)	0.76 (0.43)	0.38 (0.49)	0.42 (0.49)	0.38 (0.49)	0.35 (0.48)
Clicking sounds (0/1)	0.98 (0.13)	0.75 (0.43)	0.36 (0.49)	0.45 (0.5)	0.42 (0.49)	0.45 (0.5)
Surfacing before (0/1)	0.98 (0.13)	0.75 (0.43)	0.71 (0.45)	0.71 (0.45)	0.87 (0.33)	0.81 (0.39)
Surfacing after (0/1)	0.97 (0.17)	0.8 (0.4)	0.72 (0.45)	0.66 (0.47)	0.87 (0.34)	0.81 (0.39)

Notes: Units are shown for each data variable, except the unitless circular variance of pitch and roll, number of animals in the focal group (Group size), and presence/absence variables (marked as 0/1). Data are pooled across tag records.

associated with calves were combined in a single factor covariate. However, even in this case the improvement in AIC was small (2.2 units). We therefore concluded there was only weak support for the more complex model structures describing temporal variation in the transition probabilities, and selected the model with the lowest BIC (time-homogeneous model with four dive types) as the best-fitting model from which we made inferences about the pilot whale ethogram.

The four dive types showed distinct multivariate distributions (Appendix S1: Fig. S5). The state-dependent distributions were used to descriptively label the dive types as “Foraging” (with the highest probability of echolocation clicks), “Exploratory” (with clicking but shallower dive type), “Crowded” (with the highest average group size), and “Directed” (the most directional horizontal movement; Fig. 1). Directed dives were the most frequent dive type (40% of all dives in baseline), followed by Exploratory (33%) and Crowded dives (22%), while Foraging dives constituted the smallest proportion of dives during baseline (6%; Table 4). Foraging dives had the deepest dive depth distribution (mean dive depth 301.6 m and range 24.7–617.4 m; all statistics given in this section are computed from the observed data within each dive type). A minority ( $n = 7$ ) of the total 176 Foraging dives were shallower than 40 m. Foraging dives were also the longest in duration (7.1 min, 1.4–13.8 min), had the highest probability of clicking and presence of social sounds (0.98

for both), and had the highest probability of being preceded or followed by a surfacing (0.98 and 0.97, respectively; Table 3). After a period of NSMs, Foraging dives were the only type of dive that was more likely to be followed by another type of dive than to be repeated (Table 4).

Sea bottom depth and dive depth were correlated among the deepest Foraging dives, with the maximum correlation achieved within Foraging dives exceeding 196 m (correlation coefficient 0.93,  $N = 109$ ), and no apparent correlation for shallower Foraging dives (0.02,  $N = 55$ ). Almost all of the Foraging dives that exceeded the 196 m depth threshold (94%), and 65% of all Foraging dives, reached within 10 m of the sea bottom depth (i.e., exceeded the shallower depth contour). Foraging dives were conducted more often when animals were present in deeper habitats (Fig. 2b). Moreover, the bottom phase of Foraging dives often tracked the demersal/benthic zone (Appendix S1: Figs. S6 and S7).

Exploratory dives had the second highest probability of clicking and social sounds, but were considerably shallower than Foraging dives ( $9.9 \pm 5.0$  m, 0.86–39.1 m). Similar to Foraging dives, Exploratory dives were associated with small group sizes and slow horizontal speeds. While Exploratory dives were the most likely dive state to associate with tight group spacing ( $0.87 \pm 0.34$ ), Foraging dives were the least likely dive state to have a tight group spacing ( $0.63 \pm 0.48$ ). While Foraging dives were most

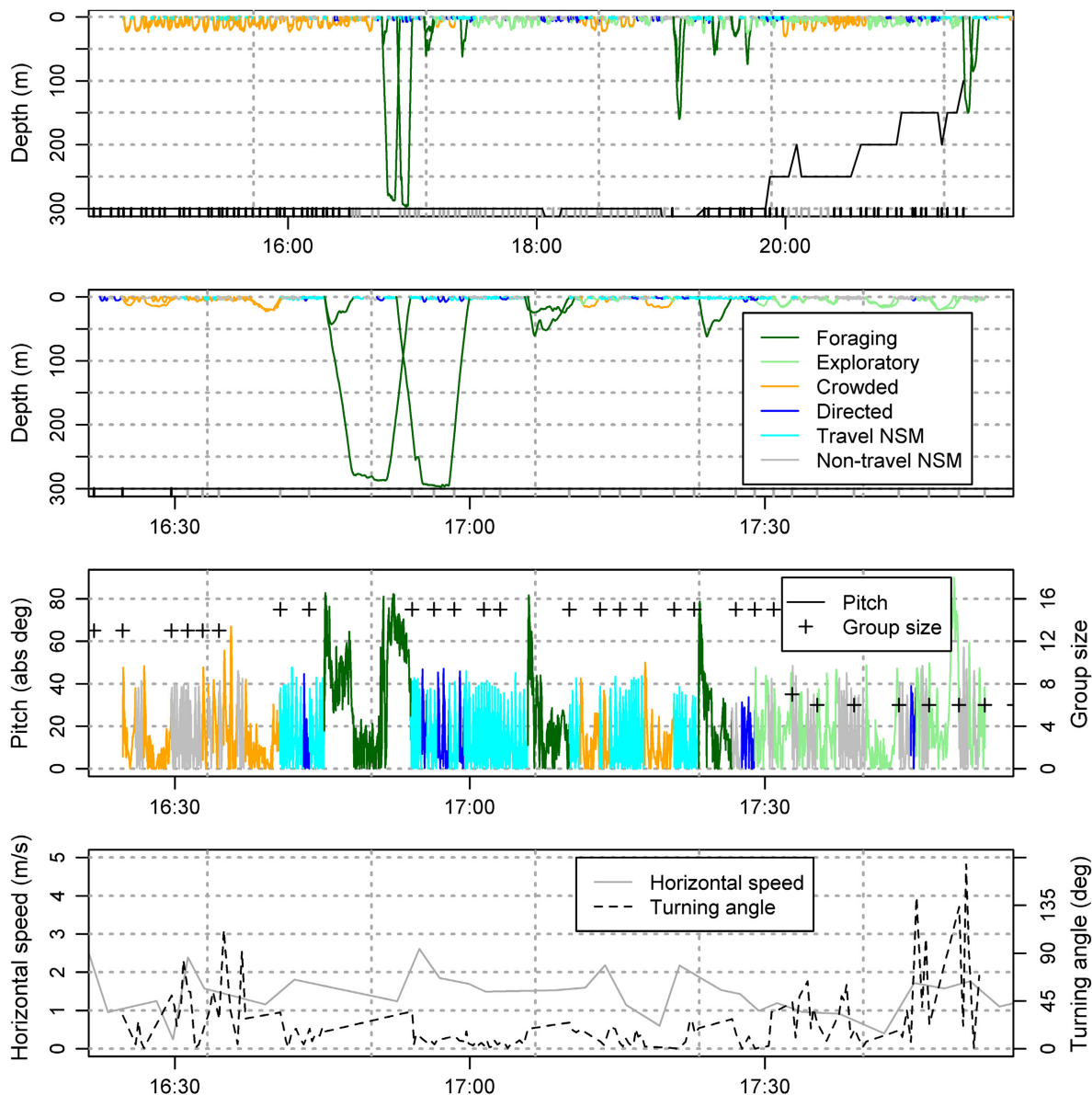


Fig. 1. Example overlapping time series of gm09\_137b and gm09\_137c. Top two panels: time-overlapping dive profiles overlaid with bathymetric contour. Rug plot shows timing of sightings, with black indicating observations of both individuals. Third panel: absolute value of pitch overlaid with visual observations of group size for the focal animal (gm09\_137b). Bottom panel: horizontal speed from visual observations, and turning angle between subsequent breaths from the tag record.

likely to transition to Exploratory dives, Exploratory dives were most likely to transition to and from Directed dives (Table 4). Directed dives were classified as the shallowest and shortest dive type with the least variation in pitch and turning angle (Table 3), and it was the only dive

type that was more frequently followed by a traveling than non-traveling NSM (Appendix S1: Table S2).

Crowded dives were estimated to have a similar duration and depth distribution to Exploratory dives (Table 3); however, they were less

Table 4. Frequency of transitions between dives in the baseline.

State	Foraging	Exploratory	Crowded	Directed
Foraging	0.299 (0.330)	0.403 (0.403)	0.065 (0.065)	0.234 (0.202)
Exploratory	0.060 (0.066)	0.738 (0.711)	0.014 (0.015)	0.189 (0.208)
Crowded	0.022 (0.014)	0.015 (0.025)	0.883 (0.871)	0.080 (0.090)
Directed	0.040 (0.033)	0.151 (0.217)	0.044 (0.055)	0.765 (0.695)
Stationary distribution	0.057 (0.059)	0.328 (0.366)	0.219 (0.222)	0.395 (0.353)

*Notes:* Values in brackets show transition probabilities estimated by the best hidden Markov model (HMM) that was fitted to all of the data (including exposures); note the close agreement with frequency of transitions during baseline. The bottom row gives the stationary distributions for both transition probability matrices. Please see Appendix S1: Table S2 for frequency of transitions including near-surface movements, and Appendix S1: Table S3 for standard errors and confidence intervals for the HMM-estimated transition probabilities.

likely to have clicking and social sounds, and group size was usually 2–3 times larger ( $20.6 \pm 8.7$  vs.  $7.1 \pm 3.4$  animals). Exploratory and Crowded dives had the highest average turning angles compared to other dives ( $>26^\circ$  vs.  $<17^\circ$ ). Crowded dives were the most likely dive type to be repeated, and similar to Exploratory dives, most frequently transitioned to Directed dives (Table 4).

#### Baseline time budgets

Individuals spent, on average, most of their time near the sea surface. In total, 24.2% of time was spent in near-surface traveling and 33.8% in other near-surface behavior during baseline (Table 5; individual average proportion of time spent in each estimated NSM). Combined, the near-surface traveling and Directed dives made up the largest proportion (35%) of the individual-average time budget. The individual-average time spent in Foraging dives was relatively small (10.3%), but also the most variable component of the time budget between baseline records (range 0–60%, coefficient of variation [CV] = 1.5). The individual-average time spent in Exploratory dives was similar to Foraging (13.3%), but unlike Foraging, occurred across all baseline records (Table 5). Individuals spent the least amount of time in the Crowded dive type (7.6%), and seven animals did not conduct any Crowded dives during baseline.

#### Synchrony of time-overlapping dive records

Both the types and timing of dives and NSMs were synchronized in portions of the time-overlapping tag records (e.g., pair gm09\_137b/c; Fig. 1). Individuals spent more time in Crowded or Exploratory dives than in Foraging dives

when the dive time between the tag records was temporally synchronized (dive time overlap  $>20\%$ ), and Foraging dives never overlapped at  $>90\%$  (Fig. 3a). Across all the pairs,  $>50\%$  of dive and surface states matched when the dive time overlap exceeded 27%, and the match increased to  $>70\%$  when the dive time overlap exceeded 70% (Fig. 3b).

The two tagged pairs A and D were scored to have more synchronous behavior than pairs B and C (Table 5). Matching dive types and NSMs constituted 55% and 47% of overlapping time in pair A and pair D, compared to 33% and 23% in B and C, respectively. The two pairs A and D also had more synchronized dives in terms of exact timing and activity ( $>75\%$  dive overlap and positive within-dive correlation of both dive depth and vertical speed), with 24.3% and 21.0% of all dives synchronized, compared to 7.7% and 5.5% for the other two pairs B and C, respectively.

#### Trade-offs in time spent foraging

The proportion of time spent in Foraging dives was selected as the response variable in the binomial GAMM as it was the dive type that was the most indicative of foraging, with the highest probability of clicking and the deepest depth distribution (*Methods*; Table 3). Twenty-one minutes was found to be sufficiently long time bin duration to remove serial correlation in the model residuals (Appendix S1: Fig. S8).

The best (lowest UBRE) model retained water depth (m), time since solar noon (h), presence/absence of sonar approaches or playbacks (SON), and order effect of sonar approaches (prev.MLFAS; Table 1). There was a good concordance between the model predicted and observed time spent in Foraging dives across tag records (Fig. 4a). We

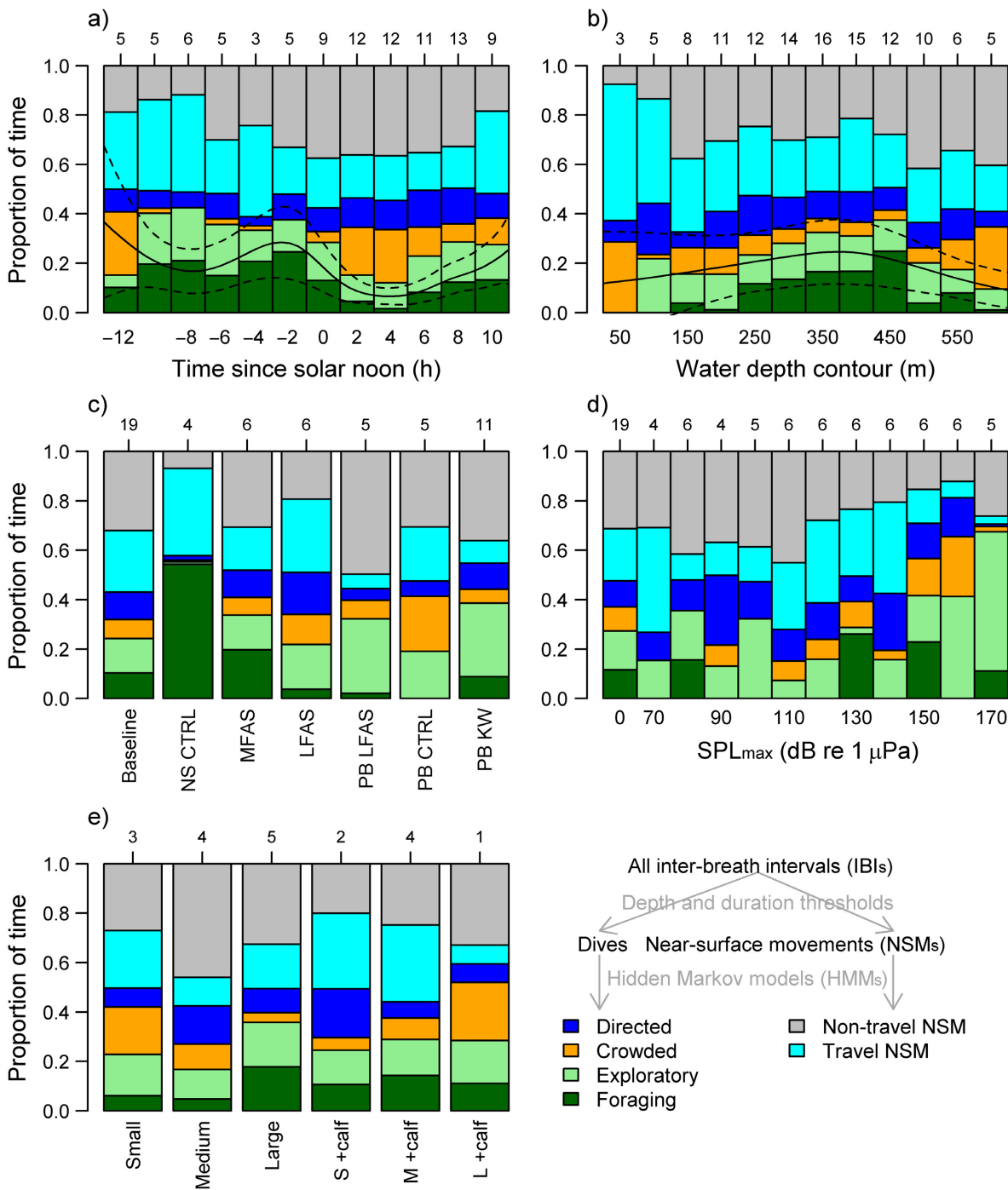


Fig. 2. Individual-average time budgets (average proportion of time per individual) as a function of environmental and individual factors during baseline and post-exposure periods (a–b, e) and during experimental exposures (c–d). Sample size (number of tag deployments) is shown for each bin at top x-axis. Panels a and b also show model predictions (see Fig. 4 for details). NS CTRL, no-sonar control approach; MFAS/LFAS, medium-/low-frequency (6–7 vs. 1–2 kHz) sonar approach; PB LFAS, near-stationary playback of sonar sounds; PB CTRL, broadband noise control playback; PB KW, playback of killer whale sounds; LFAS, low-frequency active sonar.

Table 5. Individual-average baseline time budget (%).

ID (pair ID)	Foraging	Exploratory	Crowded	Directed	NSM travel	NSM other
gm08_150c	0.0	14.7	0.0	29.0	45.3	11.1
gm08_154d	0.0	18.7	0.0	8.8	40.6	31.9
gm08_159a	7.2	8.2	0.0	24.0	43.4	17.1
gm09_137b (A)	9.0	16.5	30.0	6.9	8.5	29.1
gm09_137c (A)	7.0	15.5	22.5	8.8	23.7	22.5
gm09_138a (B)	18.4	0.8	4.8	18.2	31.5	26.3
gm09_138b (B)	1.5	5.9	2.1	26.0	30.6	33.8
gm09_156b	23.8	9.3	15.0	8.4	39.6	3.9
gm10_143a	12.0	16.1	0.0	6.1	19.3	46.5
gm10_144a	0.0	59.7	1.2	0.0	28.7	10.4
gm10_152b	3.0	13.9	13.5	8.7	39.1	21.8
gm10_157b	23.7	7.7	4.4	3.7	48.8	11.7
gm10_158d	0.0	3.7	0.0	14.1	2.1	80.1
gm13_137a	4.6	13.8	30.1	2.8	2.8	45.9
gm13_149a	60.4	0.7	0.0	1.6	8.3	29.1
gm13_169a (C)	25.2	19.0	0.0	4.8	9.3	41.8
gm13_169b (C)	0.0	22.2	0.4	7.0	7.4	63.1
gm14_180a (D)	0.0	3.8	13.4	4.6	17.7	60.5
gm14_180b (D)	0.0	2.5	6.5	21.9	13.5	55.6
Average (CV)	10.3 (1.5)	13.3 (1.0)	7.6 (1.4)	10.8 (0.8)	24.2 (0.6)	33.8 (0.2)

Notes: CV, coefficient of variation; NSM, near-surface movement. Baseline time budgets give proportion of time averaged across individual data.

found little overdispersion in the model (scale parameter estimate 1.2), indicating that the variance assumption of the binomial distribution was valid.

Foraging dives were conducted during all times of the day and in waters deeper than

150 m (Fig. 2a, b), with a model-estimated peak in proportion of time spent foraging at the 400 m maximum depth contour and a decline in deeper waters (Fig. 4d). Proportion of time spent foraging was estimated to be at its lowest 5 h after

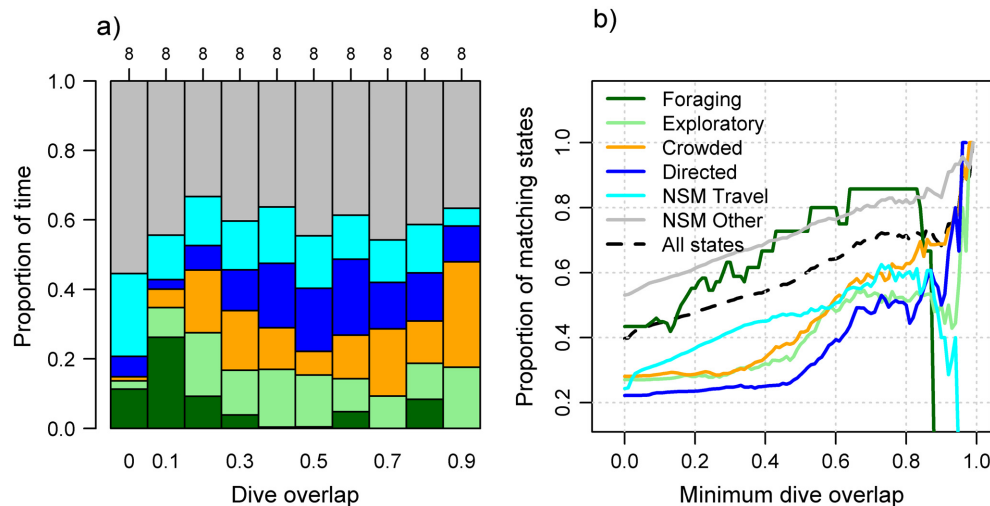


Fig. 3. Individual-average time budget (a) and proportion of matching states (b) as a function of overlapping dive time, expressed as proportion of the maximum dive time of the two tag records ( $x$ -axis). The dashed line shows the overall match between dive types when the dive overlap exceeded the value on  $x$ -axis. Solid lines show the % match for each state (color-coded), as a proportion of those states which exceeded the overlap value.

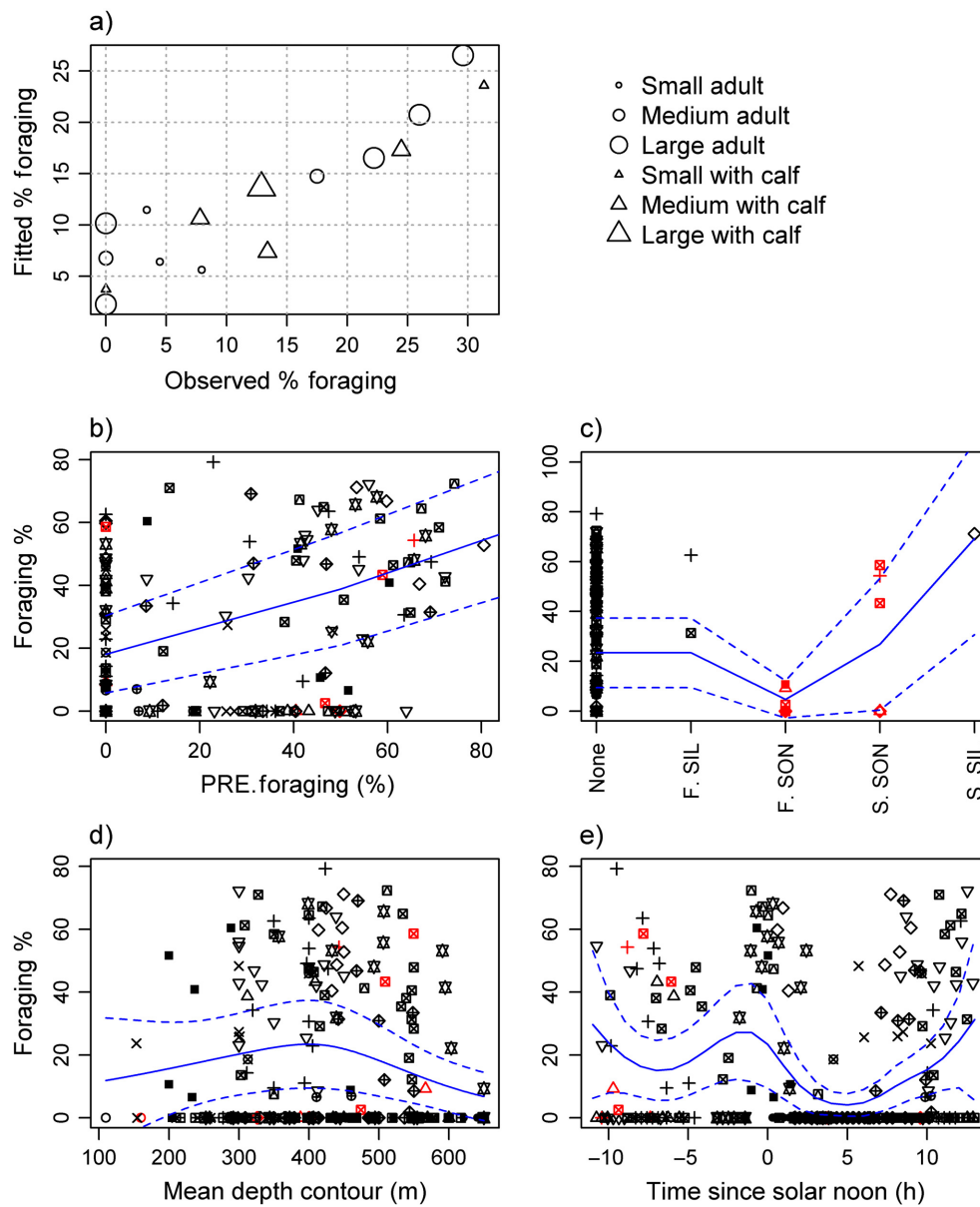


Fig. 4. Observed (symbols) and model predicted mean and 95% CI (blue lines) for percentage of time spent in Foraging dives from the best-fitting GAMM. Data and model predictions are given for each focal whale (panel a) and each 21-min time bin with symbols indicating different individuals and sonar exposure bins highlighted in red (panels b–e). Model predictions exclude random effects, with covariate values fixed (time since solar noon = 0 h; water depth = 400 m; 15% of time spent Foraging in previous 21-min time bin); observed data are shown for all covariate values and therefore do not necessarily match closely with the univariate predictions (b–e). SON, sonar exposures pooling LFAS/MFAS approaches and playbacks; F.SIL/SON, first no-sonar approach/sonar exposure; S.SIL/SON, subsequent no-sonar approaches/sonar exposures conducted after the first sonar approach; LFAS, low-frequency active sonar; MFAS, mid-frequency active sonar.



solar noon (Fig. 4e), which coincided with an increased time spent in Crowded dives (Fig. 2a) and an increase in both group size and number of animals within 200 m of the tagged whale group (Appendix S1: Fig. S7).

The individual-average time spent in Foraging dives was 10.3% during a pre-exposure baseline, 3.7% during LFAS approaches, 19.7% during MFAS approaches, 2.1% during playback of LFAS sounds, and 8.8% during playback of killer whale sounds (Fig. 2c). The best (lowest UBRE) model estimated that the ratio of time spent Foraging to non-foraging decreased by 83% (95% CI 29–96%) during sonar exposures, but increased by a factor of 7.4 (1.6–33.3; i.e., 638%) from that lower level for any subsequent no-sonar or sonar sessions (Fig. 4c). When the model selection was conducted with sonar approaches (MLFAS) and playback of sonar (PB\_SON) separately, RL\_145 was retained and MLFAS and PB\_SON excluded, but the UBRE score of this model was slightly higher (0.003, or approximately 2.2  $\Delta$ AIC units). In the model including RL\_max instead of SON, the first 20-min exposure to sonar exceeding SPL<sub>max</sub> 145 dB (SPL<sub>max</sub>: dB re 1  $\mu$ Pa) was estimated to decrease the ratio of Foraging to non-foraging time by 90% (23–99%).

## DISCUSSION

### *Ethogram and functional time budget of the pilot whale*

We identified four different dive types in long-finned pilot whales: active and mostly deep foraging dives (“Foraging”), less active and shallow dives that also contained echolocation clicks indicating foraging/exploratory behavior (“Exploratory”), non-foraging dives associated with large group sizes and lack of vocalizations (“Crowded”), and very short dives that exhibited high directionality (“Directed”). In addition, near-surface behavior could be classified to traveling and non-traveling. This ethogram could not strictly be aligned into functional behaviors (foraging, resting, traveling, and socializing), and each dive could serve multiple functions. However, we suggest that the primary function of Foraging and Exploratory dives was foraging, Crowded dives were mostly dedicated to social interactions, Directed dives and near-surface behaviors were used in horizontal travel, and the remaining near-surface behavior was resting. The

resting periods included surface intervals with little or no vertical or pitching movements to indicate separate breathing events, which most likely represented logging behavior. Overlap in the characteristics of behavior states, such as horizontal speed, indicated that multiple behavior states could associate with a function. The lack of a single foraging dive type is consistent with recent findings of Quick et al. (2017) who described several dive types representing different levels of foraging effort in short-finned pilot whales. Furthermore, social sounds and dive synchrony occurred in both foraging and non-foraging contexts, indicating that social interactions occurred across the behavioral repertoire.

On average, individuals spent most of their time near surface resting (33.8%) or transiting (35% including both near-surface travel and directed dives), or in relatively shallow (<40 m) dives (20.9%; Table 5). Both species of pilot whales have been previously reported to spend the majority of their time near surface or shallow diving (Nawojchik et al. 2003, Alves et al. 2013a, Quick et al. 2017). Resting time often has both a physiologically enforced or “conserved” component (here, a recovery period required after a breath-hold dive) and a “free” component that can be re-allocated to other behaviors that, in turn, may be more conserved in the time budget (Dunbar et al. 2009). Whether a component is free or conserved also depends upon the time-scale over which the time budgets are calculated. Here, we examined variation in within-day time budgets, unlike most studies that have concentrated on daily time budgets over longer periods of time (Marshall et al. 2012). Our analysis showed high variation in time spent foraging between baseline tag records; in contrast, time spent in non-traveling NSMs was the most conserved feature of the baseline records (lowest CV; Table 5). This may be expected due to the short tag records relative to the rate of energy acquisition in a species that carries reasonable energy reserves. For cryptic marine mammal species, estimates of the proportion of time spent at surface are useful to convert at-sea abundance estimated by visual surveys to total abundance. The relatively large and homogeneous proportion for pilot whales would imply that the abundance of this species can be estimated with relatively high precision.

Long-finned pilot whales spent 10.3% of their time in deep Foraging dives, which was relatively little time relative to other deep-diving toothed whales (Tyack et al. 2006, Watwood et al. 2006) if feeding was constrained to these dives alone. Pilot whales are thought to have high locomotion costs, as indicated by their high-performance muscle tissues, matching the high energetic content of fish and cephalopod prey that they prey upon at relatively deep depths (>200 m; Aguilar Soto et al. 2008, Spitz et al. 2012, Velten et al. 2013). Buzzes indicating prey capture attempts were recorded in greatest numbers during Foraging dives, but some were also recorded at shallow depths (<10 m) where whales might prey upon pelagic fish such as herring or cod. However, we did not include buzzes in statistical analyses as we could not use them to confirm feeding at shallow depths; further detailed analysis of their acoustic characteristics would be required to distinguish sounds produced by the tagged whale vs. other whales, the movement context of tagged whale buzz production, and to ensure buzzes are not confused with acoustically similar “rasps” that pilot whales may use in social context (Pérez et al. 2016). A more direct assessment of whether feeding occurs at shallow depths could be achieved by animal-attached video cameras or recording acoustic backscatter from prey using onboard sensors (Wisniewska et al. 2016).

#### *Benthic habitat use*

Within the range of available depths in the study area (50–700 m), individuals targeted the demersal zone or the sea bottom during the deepest parts of dives that exceeded 196 m, and spent less time foraging and more time transiting in waters with shallower depths (<200 m; Fig. 2). This confirms a general pattern of benthic diving by long-finned pilot whales in this habitat, which was previously suggested by images of the sea bottom obtained from a single whale tagged with a camera logger (Aoki et al. 2013, fig. 8 therein). There was no clear increase in foraging during particular times of day or light conditions, which may be explained by the whales’ primary reliance on echolocation and the near-continuous availability of daylight during the polar summer. However, individuals reduced time spent foraging and aggregated near the

surface in larger, more silent groups of whales in the solar afternoon. Long-finned pilot whales have been suggested to target vertically migrating prey by conducting deeper dives at night (Baird et al. 2002, Nawojchik et al. 2003, Mate et al. 2005). Our results suggest that in our study region, pilot whales instead conducted benthic or demersal dives to feed on neritic prey, which may be more accessible to them in the relatively shallow coastal area of Vestfjorden. Detailed studies on pilot whale diving behavior have concerned deep (>1000 m) pelagic habitats (long-finned: Baird et al. 2002, short-finned: Aguilar Soto et al. 2008, Jensen et al. 2011), while long-finned pilot whales are also known to inhabit shelf-edge (Nawojchik et al. 2003, Mate et al. 2005) and even shallow inshore habitats such as the fjord studied here (Nøttestad et al. 2015). Long-finned pilot whales are likely to switch between these habitats, for example, between the shelf-edge and the pelagic (Mate et al. 2005), and their seasonal movement patterns may be in relation to the location of their main cephalopod prey (Abend and Smith 1999). Concordantly, both pelagic and neritic species (e.g., *Todarodes sagittatus*) have been reported in the diet of long-finned pilot whales, with fish being more important regionally or seasonally (Desportes and Mouritsen 1988, Gannon et al. 1997). Thus, the benthic/demersal dives reported here add to the portfolio of foraging strategies that this generalist predator employs to exploit profitable food available in different habitats.

That Norwegian pilot whales frequent a coastal habitat to feed on neritic prey across multiple years (2008–2014) implies an important feeding ground in Vestfjorden. Such consistent habitat use by long-finned pilot whales, a species listed as data deficient by IUCN (International union for conservation of nature), bears conservation implications as coastal habitats are particularly vulnerable to human activities, including fishing, vessel traffic, coastal development, and run-off, which contribute to noise and chemical pollution in the neritic zone (Mann 2009).

#### *Individual differences in optimal time budgets*

We expected individuals to have different optimal time budgets and hypothesized that lactating females and larger animals would spend more time foraging if they did not match their absolute energetic requirements with increased

foraging efficiency. Energy consumption can be expected to follow an allometric function of body mass, and the energy requirement of a female pilot whale can increase by 32–63% depending on stage of lactation (Lockyer 2007). Our data were broadly consistent with this expectation, with individuals associated with calves and individuals in the largest body size category spending more than twice the time foraging than small and medium whales without calves during baseline (Fig. 2). However, neither effect was clearly supported by statistical modeling (HMM or GAMM), likely due to high contextual variability relative to number of tag records and their duration. Variation in foraging efficiency due to individual size, age, or experience may have also played a role. Larger animals are able to dive for longer periods of time due to their greater capacity to store oxygen in the body (Kooyman and Ponganis 1998) and capture larger prey, also indicated by stomach contents analyses of stranded long-finned pilot whales (Desportes and Mouritsen 1988). Furthermore, not all adults associated with calves were necessarily lactating. Long-finned pilot whale calves have been shown to associate with multiple adults, which have been suggested to provide alloparental care while foraging (Augusto et al. 2017). Future tagging studies of individuals with known size, age, body condition, and reproductive status should further elucidate the relationship between foraging efficiency and time allocation.

#### *Social foraging and behavioral synchrony*

Synchronization of behavior constitutes animals conducting the same behavior (activity synchrony) at the same time (temporal synchrony) and/or at the same place (local synchrony) and is thought to reduce risk of predation and increase social cohesion in a wide range of animal taxa (Duranton and Gaunet 2016). We found evidence of loose temporal and local synchrony of foraging dives, while activity synchrony was more apparent closer to the surface and during dive types that were less likely to involve foraging (Fig. 3). The synchronization of foraging periods rather than individual foraging dives, and association of deep foraging dives with small group sizes and loose individual spacing observed at the surface, is consistent with previous analysis on the foraging behavior of the same population of pilot

whales (Visser et al. 2014). Moreover, Foraging dives were often followed by shallower Exploratory dives (Table 4) that had similarly high probability of clicking and small group size, but unlike Foraging dives, were associated with tight group spacing at surface (Table 3). This social foraging strategy therefore appears to involve fine-scale vertical and horizontal fission of behavioral synchrony during foraging, and we were able to confirm that, at least for one tagged pair, the same individuals re-joined after such separate foraging dives (Fig. 1). Pilot whales produce social calls during and after foraging at depth (Jensen et al. 2011), which might be used to re-locate their group members after foraging dives.

A fission–fusion of behavioral synchrony such as indicated by our data implies that the cost–benefit of activity synchrony depends upon behavior state. In long-finned pilot whales, behavioral synchrony may be more expensive to perform during deep foraging dives due to increased food competition with conspecifics, and/or reduced foraging efficiency due to individual differences in energy requirements or diving capabilities (consensus costs), and locomotion costs (Aoki et al. 2013). Thus, a degree of asynchrony in foraging behavior may promote individual foraging strategies and reduce intra-specific competition for food, which in the case of pilot whales could be both scramble competition for limited benthic resources and acoustic interference during echolocation-based foraging. We hypothesize that a fission–fusion of behavioral synchrony minimizes such individual conflicts during foraging while maintaining the benefits of behavioral synchrony during non-foraging.

Re-establishing a finer degree of time and activity synchrony implies an important benefit. The costs and benefits of group living in social species are modulated by social cohesion and bonds, which may vary temporally in fission–fusion dynamics (Conradt and Roper 2000, Sueur et al. 2011). In long-finned pilot whales, synchronous breathing and diving has been suggested to function to reinforce social bonds (Senigaglia and Whitehead 2012, Aoki et al. 2013). Our results on synchronous shallow diving are in line with previous findings (Aoki et al. 2013), but further data, such as photo-identification, are required to link this behavior with preferred associates. Nevertheless, we can expect that individuals with different

optimal time budgets break their behavioral synchrony occasionally; otherwise, they would lose out on individual optimal decisions. Consistent with our hypothesis, we found that the two whales that were sighted most often within the same group at the surface (gm09\_137b/c 60% and gm14\_180a/b 34% of the focal whale sightings, compared to gm09\_138a/b, sighted together in 18% of the focal sightings, and gm13\_169a/b that were not sighted in the same group) synchronized 50% of their time budget. This was despite the fact both pairs included whales of different size classes, and while gm09\_137b was associated with a calf, gm09\_137c was not, which could lead to differences in time spent foraging. We therefore suggest that the fission–fusion of behavioral synchrony can be a strategy that allows individuals with different optimal time budgets to remain within a behaviorally cohesive group.

Potential adaptive functions of social foraging include increased availability of public information, social learning, inclusive fitness benefits, resource/anti-predatory defense, and/or benefits of group living that are not specific to foraging, such as alloparental care (Galef and Giraldeau 2001, Marshall et al. 2012). Public information can drive “local enhancement” where individuals aggregate around resources that have been discovered by others (Galef and Giraldeau 2001), and is likely to contribute to the foraging strategy of toothed whales that can acoustically eavesdrop on each other’s echolocation clicks (“dinner bell” effect) that contain information about both foraging effort (clicks) and success (terminal buzzes). Such public information may be particularly important in heterogeneous environments, where intermediate levels of resource patchiness may drive a fission–fusion strategy (Sueur et al. 2011). If pilot whales are indeed high-risk/high-benefit foragers (Aguilar Soto et al. 2008) with limited information available at the surface about the quality of the food at depth, an evolutionarily stable strategy may be supported where individuals switch between producing and scrounging public information (Galef and Giraldeau 2001). Nevertheless, benefits of social foraging in long-finned pilot whales are likely to be multiple, and not necessarily related to foraging. Indeed, besides enhanced foraging, an additional benefit of staying within a social group may be to fend off threats as shown with disturbance-specific

social responses in long-finned pilot whales (Visser et al. 2016).

#### *Trade-offs in response to disturbance*

We found that individuals traded off foraging time for time spent in shallower dives or at the surface during sonar exposures. There was a marked short-term trade-off during the first sonar exposure, with the ratio of time spent in foraging vs. time spent in other behaviors estimated to decrease by 83% (29–96%). In experiments following the first sonar approach, there was a relative increase in time spent foraging. Such an order effect might indicate habituation (i.e., increased tolerance; Bejder et al. 2009), and/or an increased tendency to avoid the source at foraging depths rather than return to near-surface behavior. The reduced foraging and concurrent increase in time spent in transiting states during sonar approaches (Fig. 2c) are consistent with previous reports of disrupted deep diving (Sivle et al. 2012) and avoidance responses to navy sonar in pilot whales in Norwegian waters (Antunes et al. 2014). Qualitative scoring of pilot whale behavioral responses (from the same dataset) indicated cessation of foraging to occur at relatively low received sound pressure levels (SPL) of 145–159 dB re: 1  $\mu$ Pa (Miller et al. 2012) compared to a high estimated SPL threshold of 170 dB re: 1  $\mu$ Pa above which 50% individuals are expected to show an avoidance response (Antunes et al. 2014). Thus, cessation of foraging may be the first of a sequence of responses where individuals return to the surface, perhaps to establish contact with their social group and/or secure faster access to air (oxygen), before engaging in a group-level and more disturbance-specific response such as horizontal avoidance or attraction (Curé et al. 2012, Visser et al. 2016). A relatively low SPL threshold for cessation of foraging may explain why the presence/absence of sonar exposures (towed sonar approaches and playbacks combined) was supported, despite differences in both source and received level.

Pilot whales have also shown a horizontal attraction response to killer whale sound playbacks, perhaps to investigate a sound source or as a mobbing response to a potential predator/food competitor (Curé et al. 2012). Despite an apparent trend in the data (Fig. 2), we found no strong statistical evidence that playback of killer whale

sounds was associated with a reduced proportion of time spent foraging. The clearer statistical support for a reduction in time spent foraging during sonar exposures, including both the towed sonar approaches and nearby playbacks, may indicate a more consistent foraging trade-off in response to the detection of sonar than killer whale sounds. In mostly solitary sperm whales, 1- to 2-kHz sonar approaches and killer whale sound playbacks were associated with a near identical reduction in time spent foraging (Isojunno et al. 2016). The discrepancy between the two deep-diving odontocete species may be explained by a different level and type (predation/competition) of perceived risk from killer whales. The perceived risk may be further modulated by conspecific behavior. A social response to disturbance in pilot whales (Curé et al. 2012, Visser et al. 2016) may allow some individuals to display a shorter duration response than the group as a whole, or even continue key activities such as foraging. Nevertheless, the weak support for exposure covariates in the HMMs and wide confidence intervals around the GAMM estimates also reflect the previously reported high inter-individual variability in behavioral changes (Miller et al. 2012), which we could partially link to time of day and water depth (Fig. 2). Thus, the apparent plasticity of individual-level time budgets at the relatively fine (within-day) temporal scale may be better explained by more direct drivers of behavior and habitat use, such as prey field. For example, Friedlaender et al. (2016) showed that echosounder data on krill density, as well as bathymetric depth, were important for predicting behavioral responses to sonar in blue whales.

Individual behavior modification to perceived costs and risks in the environment (e.g., navy sonar) may become biologically significant if individuals continued to trade fitness-enhancing behaviors (e.g., foraging time, physiologically enforced rest) for perceived safer behavior (Frid and Dill 2002). For example, killer whales reduce foraging effort in the presence of vessel traffic, which could translate to lost feeding opportunities and a substantial decrease in their energy intake (Williams et al. 2006, Lusseau et al. 2009). However, highly context-dependent time budgets may also indicate a degree of flexibility over short timescales. In the case of pilot whales, we can speculate that some of the large proportion of time spent near-surface (>60%; results herein,

Baird et al. 2002, Nawojchik et al. 2003) may represent “free” time that individuals can re-allocate without, or with less severe, biologically significant consequences to their time budgets. On the other hand, some individuals and life stages (e.g., lactating females) with higher energy requirements may be less flexible to compensate for lost foraging time. For example, high feeding rates in harbor porpoise have been suggested to increase their vulnerability to anthropogenic noise (Wisniewska et al. 2016).

#### *Methodological considerations*

Our results highlight that natural and anthropogenic drivers of individual fitness and survival are often intertwined, leading to a need for research methods that can model a multitude of individual, social, environmental, and anthropogenic processes that contribute to changes in individual behavior, reproductive success, and survival (New et al. 2014). The benefit of using multivariate mixture models and HMMs is that they can integrate multiple streams of time series data and easily allow for missing data. We showed that this approach can be used to generate ethograms from animal-borne tag data, which is a key challenge to describe the behavior of free-ranging species (Sakamoto et al. 2009). Such an approach could also be used to model, for example, the context specificity of vocalizations (Popov et al. 2017), a subject of interest in social communication. Hidden Markov models account for the time series nature of the data explicitly, but a potential drawback is the simplistic first-order Markov aspect of the model that assumes that the underlying state probability is only dependent upon the previous state. We demonstrated how this assumption can be relaxed by incorporating time-varying covariates for the transition probabilities, although in our dataset these were only weakly supported in model selection. The HMM did not include covariate effects on the time spent in different dive types or at the surface and operated on slightly different timescales (dive-by-dive vs. 21-min time bins), which may explain why covariates were retained in the GAMM for proportion of time spent foraging (based upon UBRE information criterion) while covariates were not clearly supported in the HMM (where AIC favored individual effects not supported by BIC).

The random-effect structures were also different, with the GAMM including a continuous normal random effect in contrast to the discrete random-effect structure of the HMM. On the other hand, the GAMM did not account for any uncertainty in the classification of foraging dives, but the simpler parameterization also meant that multiple covariates could be included in the same model. The two approaches can therefore give slightly different but complementary results.

The GAMM approach was sufficient to model proportion of time spent foraging, but there is room for improvement and need for best practice research modeling time budgets. We found that the timescale used to specify the number of successes and trials influenced the standard errors of the parameter estimates of the binomial GAMM. A smaller timescale increased the binomial sample size for a given bin length (e.g., 30 min/0.5 h) and thus increased the significance of covariates. We scaled the number of successes and trials to the average foraging dive duration, which allowed us to interpret the proportion of time spent foraging in a time bin with respect to an expected number of foraging dives. However, there is a need to validate this approach, and to further develop statistical packages to perform regression modeling of serially correlated time budgets (multinomial, rather than binomial, proportions) with random effects.

We report activity budgets at a fine temporal scale from a sample of 19 individuals in a specific coastal fjord habitat. Therefore, our results should be interpreted as a report of behavior within this habitat, and we do not attempt to generalize specific relationships between the whale behavior and its environment to elsewhere, such as oceanic habitats. However, the fine-scale temporal approach appears promising to apply to large datasets, which could be more readily used to identify the flexibility in activity budgets to cope with both internal (e.g., body condition) and external stressors (e.g., reduced food availability; Russell et al. 2015). For species whose populations are challenging to monitor, such as cetaceans that spend only a minority of their time at the sea surface, approaches that identify changes in time budgets could be used to understand both the mechanism and consequence of external stressors such as climate change to vulnerable populations.

## CONCLUSIONS

We quantified an ethogram and activity budget for the long-finned pilot whale and have demonstrated that coastal and benthic habitats can be important feeding grounds in this species. Fission–fusion of groups at the water surface and activity-dependent synchrony suggest a foraging strategy that minimizes individual conflict while maximizing benefits of group cohesion, such as reduction in the cost of finding food. There was a decrease in time spent foraging during the first naval sonar exposures (1–2 or 6–7 kHz), but the responses were more variable for subsequent repeat exposures. Despite previous findings of social responses to both naval sonar and killer whale sound playbacks (Visser et al. 2016), we found less evidence for significant individual-level foraging time trade-offs in response to killer whale playbacks. This is likely to be due to high plasticity of individual behavior, which we quantified here as variability in time budgets in different social and environmental contexts.

## ACKNOWLEDGMENTS

We thank 3S (Sea mammals, Sonar, Safety) ship's crew and research team members for efforts on the field data collection and access. We give special thanks to Dr. Kagari Aoki for providing the Little Leonardo tag used in this study. We would also like to thank Prof. Roland Langrock for his help and advice on fitting HMMs in R. Visual data were collected using Logger 2000, developed by the International Fund for Animal Welfare (IFAW) to promote benign and non-invasive research. We would also like to thank our sponsors, NL Ministry of Defence, NOR Ministry of Defence, U.S. Office of Naval Research (N00014-08-1-0984, N00014-10-1-0355, N00014-14-1-0390), FR Ministry of Defence (DGA; public market no. 15860052), World Wildlife Fund Norway (9E0682), and French Total Foundation and Bleustein-Blanchet Foundation. The statistical development work was supported by a separate grant from the U.S. Office of Naval Research (N00014-12-1-0204), under the project entitled Multi-study Ocean acoustics Human effects Analysis (MOCHA).

## LITERATURE CITED

Abend, A. G., and T. D. Smith. 1999. Review of distribution of the long-finned Pilot Whale (*Globicephala melas*) in the North Atlantic and Mediterranean. NOAA Technical Memorandum NMFS-NE 22, Woods Hole, Massachusetts, USA.

- Aguilar de Soto, N., M. P. Johnson, P. T. Madsen, F. Díaz, I. Domínguez, A. Brito, and P. L. Tyack. 2008. Cheetahs of the deep sea: deep foraging sprints in short-finned pilot whales off Tenerife (Canary Islands). *Journal of Animal Ecology* 77:936–947.
- Alves, F., A. Dinis, C. Ribeiro, C. Nicolau, M. Kaufmann, C. M. Fortuna, and L. Freitas. 2013a. Daytime dive characteristics from six short-finned pilot whales *Globicephala macrorhynchus* off Madeira Island. *Arquipelago* 31:1–8.
- Alves, F., S. Quérrouil, A. Dinis, C. Nicolau, C. Ribeiro, L. Freitas, M. Kaufmann, and C. Fortuna. 2013b. Population structure of short-finned pilot whales in the oceanic archipelago of Madeira based on photo-identification and genetic analyses: implications for conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* 23:758–776.
- Amos, B., C. Schlötterer, and D. Tautz. 1993. Social structure of pilot whales revealed by analytical DNA profiling. *Science* 260:670–672.
- Antunes, R., P. H. Kvasdheim, F. P. A. Lam, P. L. Tyack, L. Thomas, P. J. Wensveen, and P. J. O. Miller. 2014. High thresholds for avoidance of sonar by free-ranging long-finned pilot whales (*Globicephala melas*). *Marine Pollution Bulletin* 83:165–180.
- Aoki, K., M. Sakai, P. J. O. Miller, F. Visser, and K. Sato. 2013. Body contact and synchronous diving in long-finned pilot whales. *Behavioural Processes* 99:12–20.
- Augusto, J. F., T. R. Frasier, and H. Whitehead. 2013. Using photography to determine sex in pilot whales (*Globicephala melas*) is not possible: Males and females have similar dorsal fins. *Marine Mammal Science* 29:213–220.
- Augusto, J. F., T. R. Frasier, and H. Whitehead. 2017. Characterizing alloparental care in the pilot whale (*Globicephala melas*) population that summers off Cape Breton, Nova Scotia, Canada. *Marine Mammal Science* 33:440–456.
- Baird, R. W., J. F. Borsani, M. B. Hanson, and P. L. Tyack. 2002. Diving and night-time behavior of long-finned pilot whales in the Ligurian Sea. *Marine Ecology Progress Series* 237:301–305.
- Beale, C. M. 2007. The behavioral ecology of disturbance responses. *International Journal of Comparative Psychology* 20:111–120.
- Bejder, L., A. Samuels, H. Whitehead, H. Finn, and S. Allen. 2009. Impact assessment research: use and misuse of habituation, sensitisation and tolerance in describing wildlife responses to anthropogenic stimuli. *Marine Ecology Progress Series* 395:177–185.
- Boyd, I. L. 1997. The behavioural and physiological ecology of diving. *Trends in Ecology & Evolution* 12:213–217.
- Brown, J. S., and B. P. Kotler. 2004. Hazardous duty pay and the foraging cost of predation. *Ecology Letters* 7:999–1014.
- Conradt, L., and T. J. Roper. 2000. Activity synchrony and social cohesion: a fission-fusion model. *Proceedings. Biological Sciences* 267:2213–2218.
- Conradt, L., and T. J. Roper. 2005. Consensus decision making in animals. *Trends in Ecology and Evolution* 20:449–456.
- Côte, S. D., J. A. Schaefer, and F. Messier. 1997. Time budgets and synchrony of activities in muskoxen: the influence of sex, age, and season. *Canadian Journal of Zoology* 75:1628–1635.
- Curé, C., R. Antunes, F. Samarra, A. C. Alves, F. Visser, P. H. Kvasdheim, and P. J. O. Miller. 2012. Pilot whales attracted to killer whale sounds: acoustically-mediated interspecific interactions in Cetaceans. *PLoS ONE* 7:1–5.
- D’Amico, A., R. C. Gisiner, D. R. Ketten, J. A. Hammock, C. Johnson, P. L. Tyack, and J. Mead. 2009. Beaked whale strandings and naval exercises. *Aquatic Mammals* 35:452–472.
- de Stephanis, R., P. Verborgh, S. Pérez, R. Esteban, L. Minvielle-Sebastia, and C. Guinet. 2008. Long-term social structure of long-finned pilot whales (*Globicephala melas*) in the Strait of Gibraltar. *Acta Ethologica* 11:81–94.
- DeRuiter, S. L., R. Langrock, T. Skirbutas, J. A. Goldbogen, J. Calambokidis, A. S. Friedlaender, and B. L. Southall. 2017. A multivariate mixed hidden Markov model to analyze blue whale diving behaviour during controlled sound exposures. *Annals of Applied Statistics* 11:362–392.
- Desportes, G., and R. Mouritsen. 1988. Diet of the pilot whale, *Globicephala melas*, around the Faroe Islands. Pages 1–15 in *Intergovernmental Panel on Climate Change*, editor. ICES Council meeting (collected papers). Cambridge University Press, Cambridge, UK.
- Dunbar, R. I. M., A. H. Korstjens, and J. Lehmann. 2009. Time as an ecological constraint. *Biological Reviews* 84:413–429.
- Durantón, C., and F. Gaunet. 2016. Behavioural synchronization from an ethological perspective: overview of its adaptive value. *Adaptive Behavior* 24:181–191.
- Frid, A., and L. M. Dill. 2002. Human-caused disturbance stimuli as a form of predation risk. *Conservation Ecology* 6:11.
- Friedlaender, A. S., E. L. Hazen, J. A. Goldbogen, A. K. Stimpert, J. Calambokidis, and B. L. Southall. 2016. Prey-mediated behavioral responses of feeding blue whales in controlled sound exposure experiments. *Ecological Applications* 26:1075–1085.

- Galef, B. G., and L. A. Giraldeau. 2001. Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. *Animal Behaviour* 61:3–15.
- Gannon, D. P., A. J. Read, J. E. Craddock, K. M. Fristrup, and J. R. Nicolas. 1997. Feeding ecology of long-finned pilot whales *Globicephala melas* in the western North Atlantic. *Marine Ecology Progress Series* 148:1–10.
- Gero, S., D. Engelhaupt, L. Rendell, and H. Whitehead. 2009. Who cares? Between-group variation in alloparental caregiving in sperm whales. *Behavioral Ecology* 20:838–843.
- Gill, J. A., K. Norris, and W. J. Sutherland. 2001. Why behavioural responses may not reflect the population consequences of human disturbance. *Biological Conservation* 97:265–268.
- Harris, C. M., D. Sadykova, S. L. DeRuiter, P. L. Tyack, P. J. O. Miller, P. H. Kvasdheim, F. P. A. Lam, and L. Thomas. 2015. Dose response severity functions for acoustic disturbance in cetaceans using recurrent event survival analysis. *Ecosphere* 6:art236.
- Houston, A. I., E. Prosser, and E. Sans. 2012. The cost of disturbance: A waste of time and energy? *Oikos* 121:597–604.
- Isojunno, S., C. Curé, P. H. Kvasdheim, F. P. A. Lam, P. L. Tyack, P. J. Wensveen, and P. J. O. M. Miller. 2016. Sperm whales reduce foraging effort during exposure to 1-2 kHz sonar and killer whale sounds. *Ecological Applications* 26:77–93.
- Isojunno, S., and P. J. O. Miller. 2015. Sperm whale response to tag boat presence: Biologically informed hidden state models quantify lost feeding opportunities. *Ecosphere* 6:1–46.
- Jammalamadaka, S. R., and A. Sengupta. 2001. Topics in circular statistics. Volume 5. World Scientific Co. Pte Ltd, Singapore, Singapore.
- Jensen, F. H., J. M. Perez, M. Johnson, N. A. Soto, and P. T. Madsen. 2011. Calling under pressure: Short-finned pilot whales make social calls during deep foraging dives. *Proceedings. Biological Sciences* 278:3017–3025.
- Johnson, M. P., N. A. Soto, and P. T. Madsen. 2009. Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. *Marine Ecology Progress Series* 395: 55–73.
- Johnson, M. P., and P. L. Tyack. 2003. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering* 28:3–12.
- Kooyman, G. L., and P. J. Ponganis. 1998. The physiological basis of diving to depth: birds and mammals. *Annual Review of Physiology* 60:19–32.
- Langrock, R., T. A. Marques, R. W. Baird, and L. Thomas. 2014. Modeling the diving behavior of whales: a latent-variable approach with feedback and semi-Markovian components. *Journal of Agricultural, Biological, and Environmental Statistics* 19:82–100.
- Lockyer, C. 2007. All creatures great and smaller: a study in cetacean life history energetics. *Journal of the Marine Biological Association of the UK* 87:1035.
- Lusseau, D., D. E. Bain, R. Williams, and J. C. Smith. 2009. Vessel traffic disrupts the foraging behavior of southern resident killer whales *Orcinus orca*. *Endangered Species Research* 6:211–221.
- Mann, K. H. 2009. Ecology of coastal waters with implications for management. Second edition. John Wiley & Sons, Hoboken, New Jersey, USA.
- Marshall, H. H., A. J. Carter, J. M. Rowcliffe, and G. Cowlshaw. 2012. Linking social foraging behaviour with individual time budgets and emergent group-level phenomena. *Animal Behaviour* 84: 1295–1305.
- Mate, B. R., B. A. Lagerquist, M. Winsor, J. R. Geraci, and J. H. Prescott. 2005. Movements and dive habits of a satellite-monitored long-finned pilot whale (*Globicephala melas*) in the Northwest Atlantic. *Marine Mammal Science* 21:136–144.
- Meeus, J. H. 1991. Astronomical algorithms. Second edition. Willmann-Bell, Richmond, Virginia, USA.
- Miller, P. J. O., M. P. Johnson, and P. L. Tyack. 2004. Sperm whale behaviour indicates the use of echolocation click buzzes “creaks” in prey capture. *Proceedings. Biological Sciences* 217:2239–2247.
- Miller, P. J. O., P. H. Kvasdheim, F. A. Lam, P. J. Wensveen, R. Antunes, A. C. Alves, F. Visser, L. Kleivane, P. L. Tyack, and L. D. Sivle. 2012. The severity of behavioral changes observed during experimental exposures of killer (*Orcinus orca*), long-finned pilot (*Globicephala melas*), and sperm (*Physeter macrocephalus*) whales to naval sonar. *Aquatic Mammals* 38:362–401.
- Miller, P., et al. 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-2011-0, University of St Andrews, St Andrews, UK.
- Mintzer, V. J., D. P. Gannon, N. B. Barros, and A. J. Read. 2008. Stomach contents of mass-stranded short-finned pilot whales (*Globicephala macro-rhynchus*) from North Carolina. *Marine Mammal Science* 24:290–302.
- Nawojchik, R., D. J. St Aubin, and A. Johnson. 2003. Movements and dive behavior of two stranded,



- rehabilitated long-finned pilot whales (*Globicephala melas*) in the Northwest Atlantic. *Marine Mammal Science* 19:232–239.
- New, L. F., et al. 2014. Using short-term measures of behaviour to estimate long-term fitness of southern elephant seals. *Marine Ecology Progress Series* 496:99–108.
- Nøttestad, L., B. A. Krafft, V. Anthonypillai, M. Bernasconi, and L. Langård. 2015. Recent changes in distribution and relative abundance of cetaceans in the Norwegian Sea and their relationship with potential prey. *Frontiers in Ecology and Evolution* 2:83.
- Ottensmeyer, C. A., and H. Whitehead. 2003. Behavioural evidence for social units in long-finned pilot whales. *Canadian Journal of Zoology* 81:1327–1338.
- Pérez, J. M., F. H. Jensen, L. Rojano-Doñate, and N. Aguilar de Soto. 2016. Different modes of acoustic communication in deep-diving short-finned pilot whales (*Globicephala macrorhynchus*). *Marine Mammal Science* 33:59–79.
- Pollard, K. A., and D. T. Blumstein. 2008. Time allocation and the evolution of group size. *Animal Behaviour* 76:1683–1699.
- Popov, V., R. Langrock, S. L. DeRuiter, and F. Visser. 2017. An analysis of pilot whale vocalization activity using hidden Markov models. *Journal of the Acoustical Society of America* 141:159–171.
- Quick, N. J., S. Isojunno, D. Sadykova, M. Bowers, D. P. Nowacek, and A. J. Read. 2017. Hidden Markov models reveal complexity in the diving behaviour of short-finned pilot whales. *Scientific Reports* 7:45765.
- R Development Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rendell, L. E., and J. C. D. Gordon. 1999. Vocal response of long-finned pilot whales (*Globicephala melas*) to military sonar in the Ligurian Sea. *Marine Mammal Science* 15:198–204.
- Russell, D. J. F., B. T. McClintock, J. Matthiopoulos, P. M. Thompson, D. Thompson, P. S. Hammond, E. L. Jones, M. L. Mackenzie, S. Moss, and B. J. McConnell. 2015. Intrinsic and extrinsic drivers of activity budgets in sympatric grey and harbour seals. *Oikos* 124:1462–1472.
- Sakamoto, K. Q., K. Sato, M. Ishizuka, Y. Watanuki, A. Takahashi, F. Daunt, and S. Wanless. 2009. Can ethograms be automatically generated using body acceleration data from free-ranging birds? *PLoS ONE* 4:e5379.
- Senigaglia, V., and H. Whitehead. 2012. Synchronous breathing by pilot whales. *Marine Mammal Science* 28:213–219.
- Sih, A. 2013. Understanding variation in behavioural responses to human-induced rapid environmental change: a conceptual overview. *Animal Behaviour* 85:1077–1088.
- Sivle, L. D., P. H. Kvadsheim, A. Fahlman, F. P. A. Lam, P. L. Tyack, and P. J. O. Miller. 2012. Changes in dive behavior during naval sonar exposure in killer whales, long-finned pilot whales, and sperm whales. *Frontiers in Physiology* 3:400.
- Southall, B. L., D. P. Nowacek, P. J. O. Miller, and P. L. Tyack. 2016. Experimental field studies to measure behavioral responses of cetaceans to sonar. *Endangered Species Research* 31:293–315.
- Southall, B. L., et al. 2007. Marine mammal noise-exposure criteria: initial scientific recommendations. *Aquatic Mammals* 33:411–521.
- Spitz, J., A. W. Trites, V. Becquet, A. Brind'Amour, Y. Cherel, R. Galois, and V. Ridoux. 2012. Cost of living dictates what whales, dolphins and porpoises eat: the importance of prey quality on predator foraging strategies. *PLoS ONE* 7:e50096.
- Sueur, C., A. J. King, L. Conradt, G. Kerth, D. Lusseau, C. Mettke-Hofmann, C. M. Schaffner, L. Williams, D. Zinner, and F. Aureli. 2011. Collective decision-making and fission-fusion dynamics: a conceptual framework. *Oikos* 120:1608–1617.
- Tyack, P. L., M. Johnson, N. A. Soto, A. Sturlese, and P. T. Madsen. 2006. Extreme diving of beaked whales. *Journal of Experimental Biology* 209:4238–4253.
- Van Waerebeek, K., R. Leaper, A. N. Bakker, V. Papatavrou, D. Thiele, K. Findlay, G. P. Donovan, and P. Ensor. 2010. Odontocetes of the Southern Ocean Sanctuary. *Journal of Cetacean Research and Management* 11:315–346.
- Velten, B. P., R. M. Dillaman, S. T. Kinsey, W. A. McLellan, and D. A. Pabst. 2013. Novel locomotor muscle design in extreme deep-diving whales. *Journal of Experimental Biology* 216:1862–1871.
- Visser, F., C. Curé, P. H. Kvadsheim, F.-P. A. Lam, P. L. Tyack, and P. J. O. Miller. 2016. Disturbance-specific social responses in long-finned pilot whales, *Globicephala melas*. *Scientific Reports* 6:28641.
- Visser, F., P. J. O. Miller, R. N. Antunes, M. G. Oudejans, M. L. Mackenzie, K. Aoki, F.-P. A. Lam, P. H. Kvadsheim, J. Huisman, and P. L. Tyack. 2014. The social context of individual foraging behaviour in long-finned pilot whales (*Globicephala melas*). *Behaviour* 151:1453–1477.
- Watwood, S. L., P. J. O. Miller, M. Johnson, P. T. Madsen, and P. L. Tyack. 2006. Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). *Journal of Animal Ecology* 75:814–825.
- Whitehead, H. 1996. Babysitting, dive synchrony, and indications of alloparental care in sperm

- whales. *Behavioral Ecology and Sociobiology* 38: 237–244.
- Williams, R., D. Lusseau, and P. S. Hammond. 2006. Estimating relative energetic costs of human disturbance to killer whales (*Orcinus orca*). *Biological Conservation* 133:301–311.
- Wisniewska, D. M., M. P. Johnson, J. Teilmann, L. Rojas-Doñate, J. Shearer, S. Sveegaard, L. A. Miller, U. Siebert, and P. T. Madsen. 2016. Ultra-high foraging rates of harbor porpoises make them vulnerable to anthropogenic disturbance. *Current Biology* 26:1441–1446.
- Wood, S. N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association* 99:673–686.
- Zucchini, W., I. L. MacDonald, and R. Langrock. 2016. *Hidden Markov models for time series: an introduction using R*. Second edition. Chapman & Hall/CRC Press, Boca Raton, Florida, USA.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2044/full>