Call Usage Learning by a Beluga (*Delphinapterus leucas*) in a Categorical Matching Task

Valeria Vergara and Lance Barrett-Lennard

Vancouver Aquarium Marine Science Centre, Cetacean Research Lab

The ability to modify the structure and context of vocalizations through learning plays a key role in the social interactions of many species. The investigation of categorical matching, an aspect of contextual vocal learning, is the first step toward determining how contextual learning plays a role in the use, comprehension, and categorization of sounds in the wild. To this end, we conducted a study at the Vancouver Aquarium to test the ability of a juvenile female beluga, Qila, to respond to playbacks of two types of in-air beluga calls with vocalizations that match the category of call played (a scream, which is a vocalization type shaped over time with reinforcement and not part of this species' natural repertoire, and a pulse-train, a natural call category). We first tested Qila with random sequences of the same version of the two vocalizations with which she had been trained. Her overall success in matching all playback stimuli was above chance but not statistically so (66%). She had more difficulty matching screams (54% success) than pulse trains (80% success). We next played random sequences of six novel pulse-trains and seven novel screams, which Qila had not been trained with. She responded correctly to the set of novel stimuli of both call types in 64% of the trials, a success rate that did not differ statistically from chance. Again, she had more difficulty matching screams (55% success), relative to pulse trains (74% success). These results indicate that Qila successfully matched only pulse trains, the class that is part of this species' natural repertoire. Her poor performance on matching screams might be partly explained by a difficulty to perceive categorically a signal that lacks a function in the natural repertoire of belugas.

The essential role of vocal learning in the development of human speech has catalyzed a great deal of comparative research on phylogenetically diverse species. Marine mammals are conspicuous on the growing list of species known to be flexible vocal learners. Janik and Slater (1997, 2000) differentiate production learning, the modification of the structure of signals through auditory experience with those of other individuals, from contextual learning, which describes changes in the *comprehension* and the *usage* of pre-existing sounds precipitated by learning and experience. A longitudinal vocal development study of a beluga calf (Vergara & Barrett-Lennard, 2008) suggests that both production and usage learning play a role in the development of the beluga's natural vocal repertoire.

Vocal *comprehension* learning occurs when a receiver extracts a new meaning from a familiar signal, and vocal *usage* learning arises when animals learn to use signals already present in their repertoire in specific contexts based on experiences of other individuals' use of such signals (Janik & Slater 1997, 2000). Numerous examples illustrate the relevance of both forms of contextual learning in nature. Learning to recognize conspecifics from their signals is comprehension learning. For instance, bottlenose dolphins (*Tursiops truncatus*) learn to recognize the signature whistles of individuals with whom they share a strong social bond (Sayigh, 1992). Recognizing the meaning of alarm calls designating different predators (e.g., Seyfarth, Cheney, & Marler, 1980) is also a product of comprehension learning.

Usage learning is a more complex form of contextual learning, although Schusterman (2008) hypothesized that all mammalian infants exhibit call usage learning to some extent when they adapt their nursing calls in an interactive and synergistic fashion to the responses of their mothers. He noted, for example, that dependent Northern elephant seal pups soon learn that their own nursing calls can influence their mother's

orientation postures, facilitating access to the teat. Usage learning can thus be understood as an individual learning the social or ecological context in which a call is used (Egnor & Hauser, 2004). Some of the most compelling examples of usage learning in nature by non-human animals are in primates. Hauser (1989) used playbacks of vervet monkeys' intergroup *wrr* vocalizations to study the effect of social and auditory experience on the usage and comprehension of this vocalization by vervet infants. He found that infants in groups that had been exposed more often to intergroup encounters comprehended the meaning of the intergroup *wrr* vocalization and used adult-like exemplars of the call at an earlier age than infants in other groups.

Vocal production and usage learning can be demonstrated through conditioning experiments. The vocalizations of ecologically and phylogenetically diverse animals such as marine mammals, primates, birds, cats, dogs, and rodents have been brought under control of discriminative stimuli using operant conditioning techniques with food reinforcement (reviewed in Adret, 1993; Janik & Slater 1997; Shapiro, Slater, & Janik, 2004; Schusterman, 2008). Operant vocal behaviour is voluntary rather than reflexive and shaped by consequences (Adret, 1993; Manabe, Staddon, & Cleaveland, 1997). Thus vocal usage learning should be susceptible to operant reinforcement techniques in captivity.

Learning to produce particular calls in response to conditioning stimuli shows usage learning (Janik & Slater 2000). Shapiro et al. (2004) identified four levels of increasing complexity in the experimental demonstration of usage learning. In the first level, the animal learns to vocalize on command, responding with any call to a conditioning stimulus (a novel context). Since it can always be argued that an animal may vocalize in an arbitrary context simply because of increased arousal, the second level involves training an animal to vocalize or remain silent on command. In the third level, the animal must generalize what it learned in order to make specific vocalizations in response to different visual or acoustic cues (for example, hand signals or playbacks). Conditioning experiments successfully demonstrating this slightly more complicated third level of usage learning include common hill mynas, *Gracula religiosa* (Ginsburg, 1963), budgerigars, *Melopsittacus undulatus* (Manabe, Kawashima, & Staddon, 1995), grey seals, *Halichoerus grypus* (Shapiro et al., 2004), and Pacific walrus, *Odobenus rosmarus* (Schusterman, 2008). Prior to our experiment, three belugas at the Vancouver Aquarium had been trained to vocalize in air and to remain silent on command as part of their daily show routines. Furthermore, the three whales could produce one type of call (a scream) in response to one hand signal (a visual discriminative cue) and a different call (a pulse-train) in response to a different hand signal. Thus, these animals had already demonstrated the first three levels of usage learning.

The fourth and most complex level of usage learning, *categorical matching*, requires an animal to select and produce a vocalization from its repertoire of calls that matches the category of call presented to it (for instance, produce a whistle after hearing a whistle and a growl after a growl). A true understanding of the matching task requires not only that the subject associates a particular acoustic stimulus with a particular vocalization, but that it generalizes to a novel sample of this vocal type, and ultimately that it generalizes further to apply this not just to novel exemplars of the same category, but to novel categories, matching the stimulus regardless of the call type presented. Categorical vocal matching in marine mammals has been recently demonstrated experimentally for a gray seal, which could categorize and generalize between two call classes, "growls" and "moans" (Stansbury, de Freitas, Wu, & Janik, 2015). It is common in nature and mediates important aspects of social relationships in species in which it has been proven to exist. When a vocal response to a call occurs within a few seconds of the eliciting vocalizations and consists of a vocalization of the same type (a categorical match), the process is referred to as *antiphonal calling*. This is especially evident in signals that facilitate contact between conspecifics over long distances or in dense environments (e.g., mangabeys, *Cercocebus albigena*: Waser, 1977; Japanese macaques, *Macaca fuscata*: Sugiura & Masataka, 1995; spider monkeys, *Ateles geoffroyi*: Teixidor & Byrne, 1999; African elephants, *Loxodonta africana*: McComb, Reby,

Baker, Moss, & Sayialel, 2003). Avian examples are conspicuous in group-living species. Loud contact calls that elicit calls of the same type have been described in nearly all the species of parrot studied, and establish vocal connections between individual birds (reviewed in Bradbury, 2003). Additionally, breeding pairs of some birds engage in dueting (again, calling in an antiphonal manner) for various adaptive reasons such as maintaining reproductive synchrony and coordinating joint territorial defense (Hall, 2004).

Cetacean examples of categorical matching also abound. For example, tight temporal exchanges of matching stereotyped calls between animals that are out of visual range of one another are an important aspect of intragroup calling in killer whale pods (Miller, Shapiro, Tyack, & Solow, 2004). Since the matching calls in these exchanges do not seem to be signature calls (i.e., calls that encode individual identity), Miller and colleagues hypothesized that the exchanges may allow group members to inform each other of their positions and movement trajectories almost simultaneously, although this remains to be explored. We suggest that another plausible explanation is that by responding with a categorically-matched call in a tight temporal sequence a whale is confirming in an unambiguous, easily recognized manner that it has heard the first caller. Wild bottlenose dolphins use their learned signature whistles in matching interactions, presumably to address associates, whereby an individual responds to a whistle of a conspecific by emitting the same whistle type (Janik, 2000). Sperm whales exchange matching codas (patterned sequences of clicks) seemingly to reinforce social bonds (Schulz, Whitehead, Gero, & Rendell, 2008). Belugas engage in antiphonal exchanges of contact calls presumably to maintain group cohesion and mother-calf contact (Vergara, Michaud, & Barrett-Lennard, 2010),

Belugas are well known for their plasticity in their vocal production. According to Eaton (1979), a 15-year-old beluga at the Vancouver Aquarium produced utterances that resembled human speech. Another 9-year old male beluga, held at San Diego Bay, imitated the sound of human conversations, and did so with his head out of the water (Ridgway, Carder, & Jeffries, 1985). This behavior occurred frequently enough that the trainers were able to solicit it with a fish reward, so that his vocalizations could more easily be recorded and analyzed. Murayama, Iijima, Kastumata, and Arai (2014) also demonstrated, experimentally, an aptitude for imitation in a male beluga, who could imitate playbacks of its own sounds, computer-generated artificial sounds, and human vocal sounds.

The anecdotal and experimental evidence of vocal imitation in belugas (Eaton, 1979; Murayama et al., 2014; Ridgway et al., 1985) and the indication that at least part of their large repertoire might be learned (Vergara & Barrett-Lennard, 2008) make belugas a compelling species for studying learning in the vocal domain. Further, since antiphonal calling plays a role in beluga communication, it seems reasonable to expect that they can categorize incoming calls and respond with calls of the same type. The objective of this study was to investigate this mechanism, categorical matching, under controlled conditions by testing the ability of belugas to respond to playbacks of two categories of beluga calls with matching vocalizations.

Method

Subject, Training Regime, and General Procedure

This study consisted of two phases. During Phase 1, Qila, a 10-year-old female at the time of the study, was trained to vocalize after acoustic playbacks, while Phase 2 was the strict categorical matching phase of the experiment. Qila was housed in an outdoor pool (18 m x 29 m; depth 6 m) connected to an adjacent 3 m deep 3 m x 3 m smaller medical holding pool. She shared this enclosure with her 18-year-old mother Aurora, a 2-year-old male calf, two older, unrelated females, and an unrelated adult male.

Qila was trained one to two times a day, three days a week, by one of the marine mammal trainers at the Vancouver Aquarium. Sessions took place either in the smaller medical holding pool, or at a training station at the corner of the larger pool. The experimenter was present at every session positioned behind the trainer, and videotaped each session for later analysis.

During training procedures, the trainer identified a correct response by immediately blowing a whistle (a "bridge" or secondary reinforcer), which was followed by a fish reward as a primary reinforcer (often accompanied by praising and petting). A Least Reinforcing Stimulus was used for incorrect responses. This is a neutral response so that the animal is neither punished nor rewarded, withholding overt interaction for 5-10 s (and not using the whistle), followed by another opportunity to earn a reward.

Playback Stimuli

Qila was trained with the two sound types that she regularly produced after hand signals during public shows: aurally and spectrographically distinct "screams" and "pulse-trains" (Figure 1). Screams are series of short burst pulsed sounds produced in rapid succession, each burst pulsed sound lasting 0.4 ± 0.1 seconds with a mean pulse repetition rate of 513 pulses/second ($\pm$ 79, n = 12). Most energy is below 16 kHz, with dominant frequency below 4 kHz. Pulse-trains are longer, averaging 2.02 ± 0.7 s and having a pulse repetition rate of 118 ± 4 pulses/second (n = 12). Most energy is below 16 kHz, with a dominant frequency below 5.5 kHz.
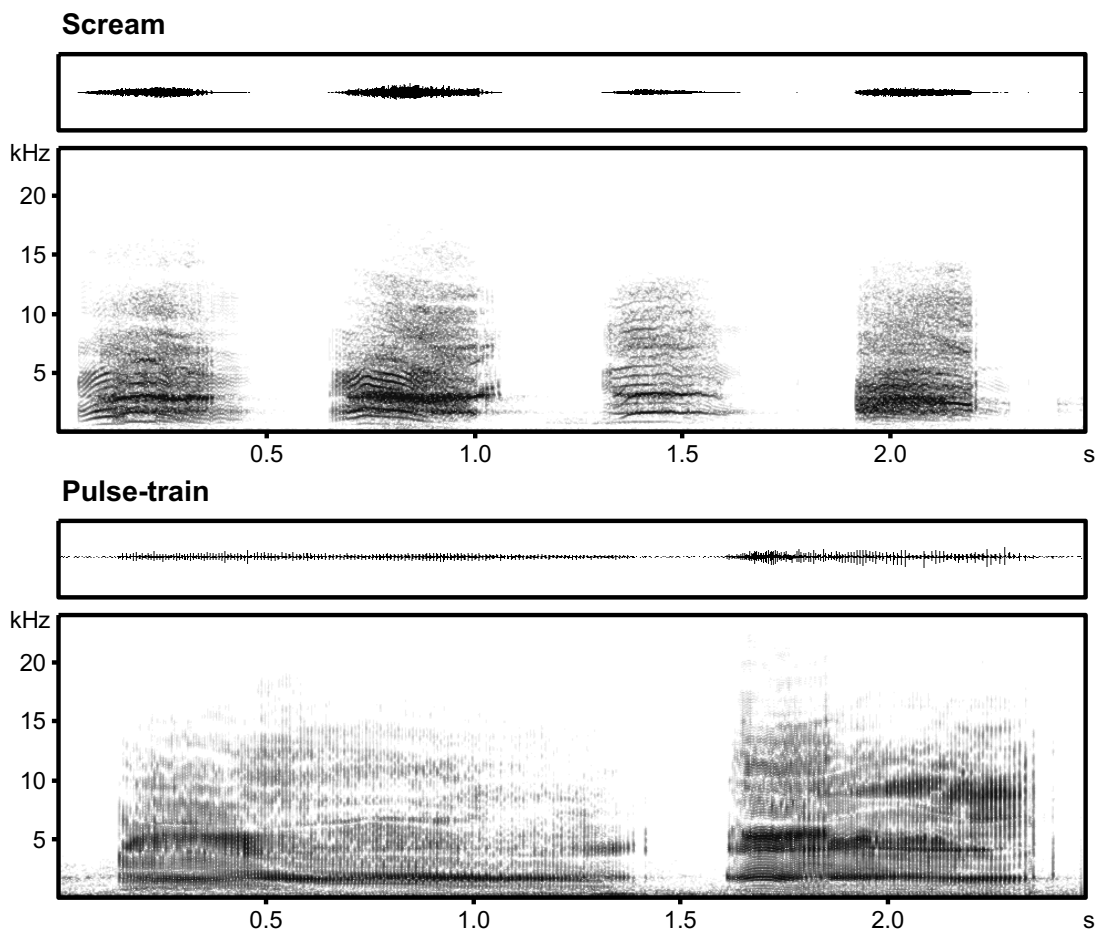

Figure 1. **Waveforms and spectrograms of the two signal types used in training.** The scream and the pulse-train (Fast Fourier Transform length: 512 points, window: hamming, frame size: 100%, overlap: 75%, frequency resolution: 86 Hz, time resolution: 2.9 ms).

In the captive setting, these two call types functioned differently. Screams were produced only during shows and during the study described here and only in air (they were never heard underwater). They were not in these belugas' natural repertoire; but were instead *shaped* entirely by trainers through selective reward. Pulse-trains, on the other hand, were used spontaneously both underwater and in-air in contexts other than shows, and were *captured* by bridging and reinforcing their production, to be used on command during shows.

The recordings for the playbacks were made from Qila's vocal responses to the two hand signals that elicit these two sound categories, using a Sony TCD-D100 Digital Audio Tape recorder and a Logitech microphone (this system had a frequency response of 0.02 Hz – 22 KHz $\pm$ 1 dB). Playbacks were broadcasted in air using a Sony digital audio player CFD-S20CP with a frequency response of 20 Hz-20 kHz (+1/-2dB). The audio player was positioned on the pool ledge, about 1 m from the rostrum of the whale. Playbacks were standardized in amplitude and were of similar duration: 2.5 to 3.5 s per playback, a length of time that could fit a total of 2 individual pulse trains or 4 individual screams. Since the latter were much shorter vocalizations, always uttered in a tight series of short bursts, we attempted to standardize the length of time rather than the number of elements.

Phase 1: Training Procedure

From October 25, 2004 until March 08, 2005, Qila was trained to vocalize after the playbacks of her own calls, thus shifting the context of these two vocalizations from visual stimuli (hand signals) to acoustic stimuli (playbacks). The trainer was the same person through the duration of the study.

Qila never responded with any other call type; the only ambiguous responses occurred when she occasionally offered both signals following a playback stimulus, which was considered an incorrect response. Other than this occasional dual response, Qila adhered strictly to pulse-trains or screams, so that assessing the correctness of the response was unproblematic.

Each session began with Qila stationed by the pool ledge, with her head out of the water, facing the trainer. The trainer ensured that she had Qila's attention by bridging her for touching the trainers' hand. Qila then kept her head vertically above the water, awaiting the next signal. The experimenter was positioned behind the trainer and off to the side, handling the playback equipment and the video camera. Since training sessions were as short as 5 min during busy Aquarium visitor times, and as long as 20 min at other times, the number of training trials per session varied between 5 and 20. The procedure consisted of three stages, described in the following sections.

Stage 1: Training to match scream playbacks. Qila was taught to produce screams after hearing screams. Initially, the experimenter played a scream, the trainer made the corresponding hand signal, then rewarded a correct vocal response. The hand signal was gradually phased out (using it only on some trials, judged by the trainer) so that the Qila would learn to attend only to the acoustic stimuli, in essence substituting one discriminative stimulus (hand cue) for another (acoustic cue). Qila required 12 training sessions (80 trials) to first produce a scream after hearing a scream playback, without relying on the hand cue.

Stage 2: Training to match pulse-train playbacks. The pulse-train playback was introduced to Qila and the same procedure repeated, i.e., the hand signal indicating "pulse train" was gradually phased out as Qila learned to respond to the pulse-train playback with a pulse-train vocalization. Qila learned this stage more rapidly than Stage 1, requiring 8 sessions (55 trials) to first produce a pulsetrain after a pulse-train playback, without relying on hand cues.

Stage 3: Training to match both playback types. This stage consisted of training Qila to respond to both signal types in the same session, matching the appropriate sound category. Up until this point, playback sequences were not randomized, but were selected by the trainer, depending on her performance. For instance, if Qila responded incorrectly to a scream, the trainer asked the experimenter to play another scream to give Qila the chance to match it and receive a reward. After session 65 (45 sessions after we started playing both types in the same session) the trainer considered Qila ready for the randomized training trials of the categorical matching experiment.

Phase 2: Categorical Matching Experiment

From March 15, 2005 until May 10, 2005 we tested Qila's categorical matching ability, adapting some of the methodology from Shapiro et al.'s (2004) study of usage learning in captive grey seals. Experimental sessions were 10-20 min long, to fit them within the institution's busy schedule, allowing for 10 to 25 trials per session. The fact that the sequences were generated at random and that the trainer determined when to end a session (based on her schedule and availability) meant that the total number of exemplars presented for each category was not the same across trials. The experimental protocol consisted of the following three trial types:

Trained-set trials. On March 15, 2005 we began presenting random sequences of the same two vocalizations with which Qila had been trained. Sequences were generated with a random number generator before every session, allowing for no more than 3 vocalizations of the same type to be played consecutively. We used exemplars of the scream and pulse-train playback with which she had been trained (3 of each). Each track on the CD played only one vocalization exemplar, i.e., tracks 1 to 3 for scream exemplars # 1 to # 3, and tracks 4 to 6 for pulse-train exemplars #1 to #3. The number of trials varied among sessions, for a total of 180 trials in 12 training sessions.

Transfer-set trials. The purpose of this phase was to assess Qila's ability to generalize the task to novel stimuli. On April 14, 2005, we began testing Qila's responses to new versions of the two call categories in a series of 10 sessions (this number of transfer trial sessions to be run was predetermined), running at least 10 trials per session, for a total of 185 trials. We used seven novel exemplars of Qila's scream vocalizations and six novel exemplars of her pulse-train vocalizations not previously used in training. We generated a random sequence for each session playing no more than three vocalizations of the same type consecutively.

Calibration trials. Prior to this set of experiments, Qila never had to repeat her vocal responses more than two or three consecutive times, because she was usually asked to produce just one vocalization of each of the two trained types during shows. We thus considered the possibility that Qila would lose motivation when the same task was asked of her 12 or more consecutive times, session after session, as in the training and transfer trials above, and that this might affect her success rate. To help distill the motivation factor from any real difficulty that she might have had in the above experiments, we ran a series of *calibration* trials testing Qila with the two types of hand signals that she responded to correctly during the shows. If her motivation was decreasing, we would expect her performance, usually perfect with just two trials per show, to drop when she was asked to repeat her task on 12 or more consecutive trials.

Starting in February 16, 2006, we conducted 10 sessions, each consisting of 12 to 16 hand signals (trials), for a total of 134 trials. As with the playback trials, we generated a random sequence for each session allowing for no more than three hand signals of the same type to be offered consecutively.

Statistical Analyses

We used generalized linear mixed effects models (GLMM), implemented with the lmer function in the 'lme4' package (Bates & Sarkar, 2006) of the *R System for Statistical Computing* (R Development Core Team, 2010). GLMMs provide a more flexible approach than strictly fixed effect methods (such as logistic regression or contingency tests) for analyzing binary data that involve random effects (Bolker et al., 2008). This approach allowed for different sources of variation in whether Qila succeeded or failed at matching the vocalization type. Playback type was treated as a fixed effect. Session (i.e., the day of the trial) was treated as a categorical random effect, allowing for Qila to have good days and bad days. To allow for the possibility that Qila learned over time, which would be demonstrated by a consistent increase in success rate over the 10-12 sessions of an experiment, we considered a model that treated session as a continuous fixed effect. Since the number of trials per session varied between sessions, we also considered a model that included the number of trials per session as a factor.

Results

For all three trial types (trained set, transfer set, and calibration), we accounted for the possibility that Qila learned over time and that the number of trials per session may have influenced Qila's performance. A model selection process based on the AIC (akaike information criterion) to find the model of the best fit indicated that the addition of these variables had little effect on model specification. We thus dropped these covariates from the model to increase the precision of the remaining parameter estimates and looked at whether Qila's overall success rate, and her success rate for each call type, differed statistically from chance. We accounted for random variation between sessions, but the variance components for the random effect of session were very small, near zero, which is indicative of sparse data.

Trained-Set Trials

Figure 2 illustrates Qila's performance on the trials that used the stimuli with which she had been trained. Qila's overall success in matching all playback stimuli was 66%, slightly above chance but not

significantly so, $Z = 1.068$, $n = 180$, $p = 0.285$. However, this might be because she had more difficulty matching screams (54% success) than pulse trains (80% success). Indeed, when playback type is included as a covariate, Qila matched the pulse-trains more often than predicted by chance, $Z = 3.268$, $n = 85$, $p < 0.01$, but not the screams, $Z = -1.713$, $n = 95$, $p = 0.087$. Her success rate matching pulse trains and screams was statistically different (i.e., in a model testing for the difference in success rate between call types, $Z = -3.639$, $n = 180$, $p < 0.001$.
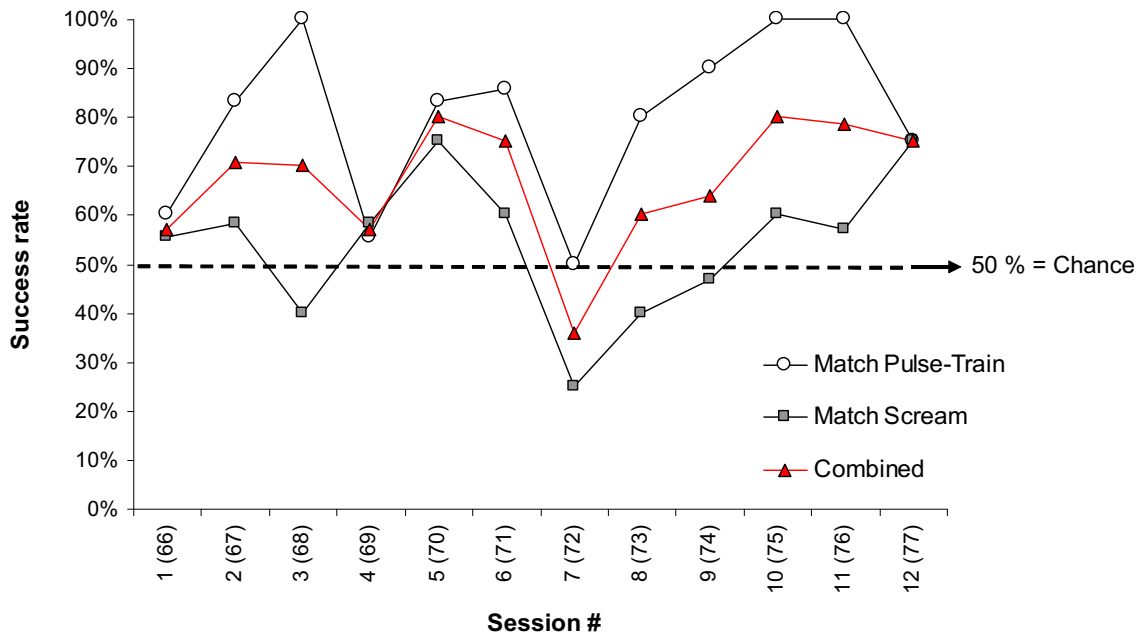

Figure 2. **Proportion of categorical matches by a female beluga, Qila, at the Vancouver Aquarium.** Matching was in response to the scream (grey squares) and pulse-train (white circles) playbacks with which she was trained, and overall success rate for both playback types combined (red triangles).The numbers in brackets on the x axis indicate the number of sessions from the beginning of Qila's training. Session # 66 was the first session where we began randomizing the playbacks to test Qila on the training trials.

Some of the responses were classified as incorrect because Qila offered both vocalization categories (shifting from one to the other without a break, which she did only when we played screams). She replied to 10 out of 95 scream playbacks with both vocalizations, always producing the pulse-train ahead of the scream.

It is evident from Figure 2 that some sessions were better than others. In the 5th and 12th (last) session, Qila performed similarly well matching pulse trains and screams, for an overall success rate of 80% (5th session) and 75% (last session), but her performance oscillated over the interim sessions.

Transfer-Set Trials

Figure 3 illustrates Qila's performance on the transfer trials. Overall, Qila responded correctly to the set of novel stimuli of both call types in 64% of the trials, a success rate that did not differ statistically from

chance, $Z = 0.431$, $n = 185$, $p = 0.666$. As in the training trials, this might be because she had difficulty matching screams (55% success), relative to pulse trains (74% success). Qila matched novel pulse-trains more often than predicted by chance, $Z = 2.509$, $n = 88$, $p = 0.012$, but not novel screams, $Z = -1.778$, $n = 97$, $p = 0.075$. The difference in success rate between pulse train and scream playbacks was statistically significant, $Z = -3.070$, $n = 185$, $p = 0.002$. Figure 3 indicates some clear variation among sessions. On the last two sessions of the transfer trials Qila matched screams and pulse-trains with approximately equal success, for an overall success of 75% (40 trials).
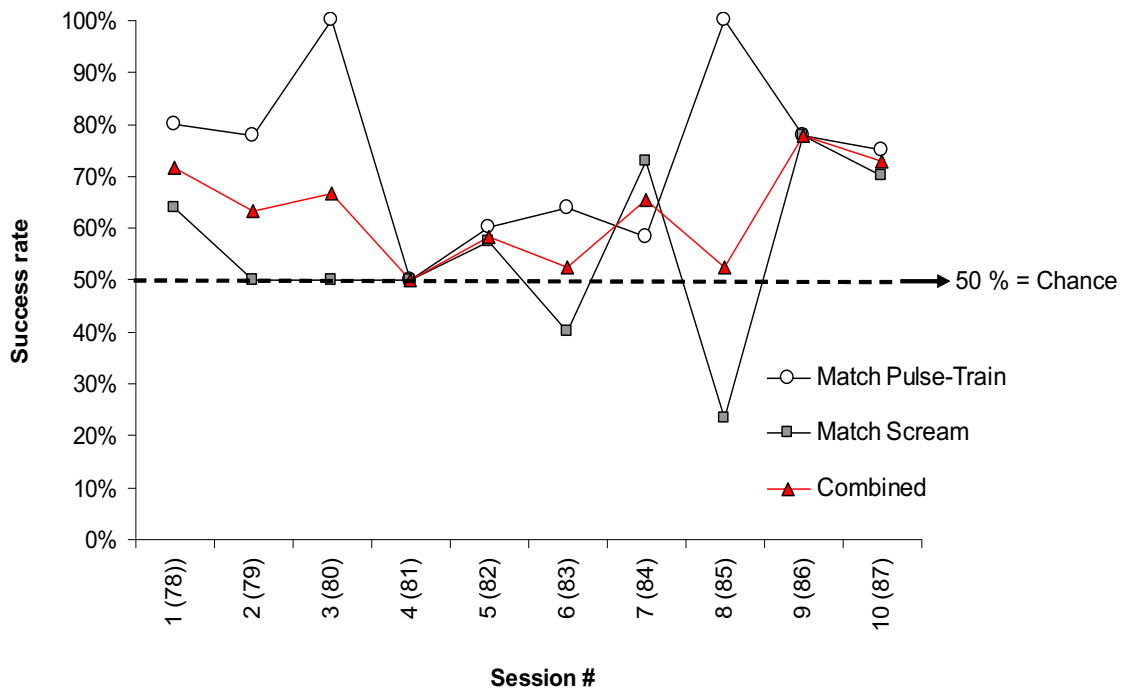

Figure 3. **Proportion of categorical matches by a female beluga, Qila, at the Vancouver Aquarium.**
Matching was in response to seven novel exemplars of the scream playbacks (grey squares) and six novel exemplars of the pulse-train playbacks (white circles), and overall success rate for both playback types combined (red triangles). The numbers in brackets on the x axis indicate the number of sessions from the beginning of Qila's training.

As during the trained-set trials, some of the responses were classified as incorrect because Qila offered both vocalization categories. Qila responded to 8 out of 97 screams with both types, always producing the pulse-train first, and to 1 out of 88 pulse trains with both vocalization types, offering the scream first.

Calibration Trials

Excluding five non-responses (i.e., Qila remained silent following two of the scream hand signals and three of the pulse-train hand signals), Qila responded to the two visual stimuli (hand signals) with the correct vocalization type 91% of the trials, significantly better than chance, $Z = 5.894$, $n = 130$, $p < 0.001$ (Figure 4).
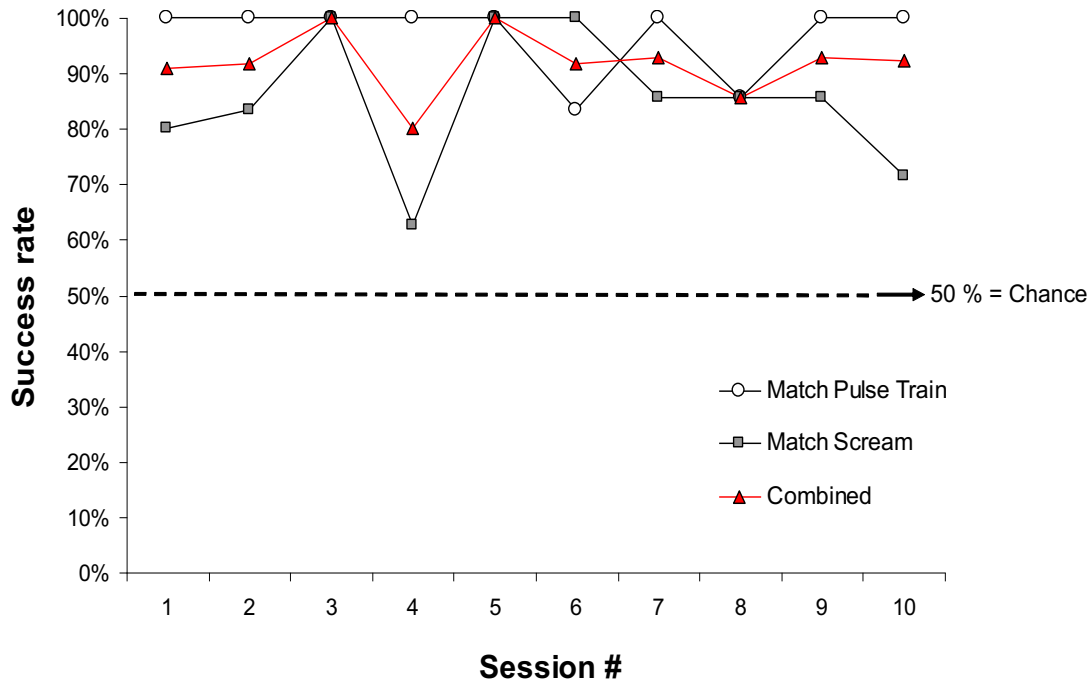
Figure 4. **Proportion of correct vocal responses by a female beluga, Qila, at the Vancouver Aquarium.** Vocal responses were to the scream hand signals (grey squares), the pulse-train hand signals (white circles), and to both signal types combined (red triangles).

Although she performed well above chance in response to each hand signal type (98% for pulse trains: $Z = 4.107$, $n = 65$, $p < 0.001$; 85% for screams: $Z = 3.504$, $n = 65$, $p < 0.001$), there was a statistically significant difference in performance between call types, $Z = -2.192$, $n = 130$, $p = 0.028$.

Discussion

The ability of a female beluga, Qila, during the training phase of this study (Phase 1) to emit different vocalizations and to remain silent on command, and the rapidity with which she switched the context of these vocalizations from visual to acoustic stimuli during the training process, point to flexibility in the vocal domain and to mastery of the first three levels of usage learning described by Shapiro et al. (2004). Qila's performance in the categorical matching experiment (Phase 2) strengthens the notion that belugas are, at the very least, proficient in the first three levels – since she had to learn to produce elements of her own vocal repertoire in this very specific playback context. Overall, however, the results do not warrant a firm conclusion that Qila could match, and discriminate between, the playback stimuli. She required what seemed like a large number of training sessions (although what constitutes a "large number" is debatable, as we argue later) to achieve a level of proficiency that was not statistically better than chance. This poor overall success appears to be related

to her difficulty matching screams. She performed above chance for pulse-trains but not screams, both when responding to call exemplars with which she had been trained, and to novel exemplars of the same call types.

It appears that Qila's success in matching pulse trains, and in generalizing this response to new auditory stimuli of the same class, indicates a true categorical matching ability for this particular vocalization type. Yet this should be interpreted with caution, as Qila could have favored one call type in her responses regardless of which playback stimulus was presented, which would explain her success matching pulse-trains. In addition, matching of either call type in novel playback sessions was quite inconsistent. For a convincing demonstration that Qila understood the premise of the categorical matching task, she would have had to perform above chance at matching *each* type. Further, the next step would require presenting Qila with an altogether different vocal class – a third category – to determine if she would transfer her knowledge of matching screams and pulse trains by producing the same category of call as the new stimulus.

Given the importance of auditory categorization in communication systems (Stansbury et al., 2015), and the anecdotal and experimental evidence of vocal imitation and plasticity in this species (Eaton, 1974; Murayama et al., 2014; Ridgway et al.,1985), these results were unexpected. The following sections evaluate the various possible explanations for why this particular beluga showed difficulty mastering this vocal task.

Perception of Vocal Categories

The question of whether categories of sounds distinct to researchers are distinct to animals arises repeatedly in the literature (Baugh, Akre, & Ryan., 2008; Fischer, 1998; Nelson & Marler, 1989; Owren, 1990; Sinnott, 1987; Snowdon & Pola, 1978). A handful of studies have found that humans differ from the studied species in the partitioning of the species-typical vocal repertoire (e.g., Snowdon & Pola, 1978, for trill vocalizations of pigmy marmosets, *Cebuella pygmaea*; Cheney & Seyfarth, 1982, for vervet monkey grunts; Weary, 1989, for songs of great tits, *Parus major*). For example, Cheney and Seyfarth (1982) found, through playbacks, that the grunt vocalization of vervet monkeys, which humans perceive as one call type, is perceived by the monkeys as four different call types that elicit different responses.

We currently lack an adequate understanding of what components of the vocalizations are perceptually important to belugas. Differences in perceptual boundaries between cetaceans and humans could partly explain Qila's difficulty. The two pulsed signals we used are clearly distinct to humans, based on their pulse repetition rate. Above a pulse repetition rate of around 300 pulses per second, humans can no longer discriminate the pulses as separate from one another (Au & Hastings, 2009), and begin to hear continuous tonal-like sounds, which sound subjectively like screams, squeals, or moans (Caldwell & Caldwell, 1967; Murray, Mercado, & Roitblat, 1998). Screams in this study were well above this threshold, whereas pulse-trains were well below it (see methods), which explains why we perceive them as distinct. Cetaceans, on the other hand, have a much finer temporal resolution of sounds. Belugas in particular have extremely refined bioacoustic abilities suited for the Arctic environment (Turl & Penner, 1989; Turl, Penner, & Au, 1987; Turl, Skaar, & Au, 1991). Qila may have had problems with the categorical perception task if these particular signals represent, to a beluga, points on a continuum rather than discrete categories. Pulse repetition rate may not be a perceptually salient feature of these two vocalizations.

To determine the perceptual validity of the two vocal classes in our study, we would need to disentangle the categorical perception task from the vocal matching task. If Qila had difficulty understanding the vocal matching task, but had no trouble distinguishing the categories of sounds presented to her, she might

achieve proficiency in the categorical perception task if she was asked to indicate her understanding with a response other than a matching vocalization. Training her to touch different objects with her melon upon hearing the different vocal categories might work. An experiment of this sort was conducted by Le Prell, Hauser, and Moody (2002), who trained rhesus monkeys to either maintain contact with, or to release, a metal cylinder during presentation of two rhesus scream categories, a tonal scream vs. an arched scream. The monkeys learned to classify initial exemplars correctly, but did not generalize the response to novel exemplars of the two categories, indicating that these classes were not perceptually discrete. The variable responses to novel tonal and arched screams suggested that the monkeys did not perceive these scream classes defined by the researchers as categorically distinct.

Categorical Perception in Relation to Function

There is evidence that animals categorize signals based on their meaning under natural circumstances (e.g., alarm calls of vervet monkeys: Seyfarth et al., 1980; lactating calls of house mice, *Mus musculus*: Ehret & Haack, 1981; vervet monkey grunts: Cheney & Seyfarth, 1982; rhesus monkey, (*Macaca mulatta*) screams: Gouzoules, Gouzoules, & Marler, 1984; food calls of rhesus monkeys: Hauser, 1998; mating calls of túngara frogs, *Physalaemus pustulosus*: Baugh et al., 2008). Baugh et al.'s (2008) study is a good example. Closely related species of túngara frogs have mating calls named "whines" that elicit the approach of females. Baugh et al. tested female *Physalaemus pustulosus* responses to a series of synthetic calls that varied continuously in several dimensions between the conspecific whine and the call of an allopatric heterospecific. Females discriminated the calls categorically into either conspecific or not conspecific. When they were presented with pairs of calls that differed by the same magnitude, females discriminated between those in different categories (a conspecific and a heterospecific call) but not between those that belonged to the same category (both conspecific calls).

Quite relevant to this discussion is the growing body of literature on phoneme recognition in humans, which stresses that knowledge of word meaning facilitates the perceptual discrimination of speech sounds (e.g., Davidson, Shaw, & Adams, 2007; Kazanina, Phillips, & Idsardi, 2006; Pfordresher & Brown, 2009). Research on cross-language speech perception is particularly illuminating. Kazanina et al (2006) examined the categorical perception of [d] and [t] sounds by Russian and Korean speakers. Only in Russian are these two sounds used to encode word meaning. Predictably, Russians could distinguish [d] from [t] easily, whereas discriminating the two sound categories was difficult for Koreans, as indicated both by behavioral tasks and electrophysiological measures of brain responses.

Along the same lines, although with an innovative angle, Pfordresher and Brown (2009) compared musical pitch perception in tone language speakers (Mandarin, Cantonese, and Vietnamese) to that of native English speakers. Individuals whose native language is a tone language, in which pitch conveys information about word meaning, are better able to imitate (through singing) and perceive pitch differences in music than speakers of English (which does not employ pitch to convey word meaning).

Collectively, the evidence reviewed here sheds light on Qila's difficulty in matching screams. This call type is not used by belugas for communication; at least not by this study group. It is a vocalization that was shaped by trainers in the older animals (Aurora and Imaq) through selectively rewarding successive approximations to the final, desired call type (a scream). Qila learned screams by mimicking her mother during training sessions; once she began volunteering scream-like sounds, their production was appropriately reinforced and further shaped by the trainers. Prior to the current study, Qila produced this call type only in air

(never underwater) and only during shows, strictly upon presentation of the discriminative stimulus (the hand signal representing this call type). She was never heard to produce this airborne sound at any other time. The pulse-train, on the other hand, is used spontaneously underwater by all the individuals in contexts other than shows. It is not surprising, then, that Qila was successful only at matching pulse-trains—part of her natural repertoire—but not screams, in both the training and the transfer trials. It is possible that Qila had difficulty perceiving screams categorically because they play no functional role in this species' natural communication system. Interestingly, the difference between her performance matching pulse-trains and matching screams was supported statistically in all three experiments. Even during the successful calibration trials, in which Qila responded to both hand signal types better than chance, she performed better for pulse-trains than for screams. These results point to the notion that Qila could not deal efficiently with a call type that is foreign to her natural vocal repertoire in tasks that involve the categorical perception and matching of such sound. This is, however, surprising, given the evidence of belugas as proficient vocal mimics of human signals that clearly do not play a role in their natural repertoire (Eaton, 1975; Murayama et al., 2014; Ridgway et al., 1985).

Motivation

Another idea worth considering is the importance of an animal's motivation on the success rate of a task. Pepperberg (2002), when training parrots in vocal learning tasks, noted that concurrent work on several tasks was crucial because the birds would become restless in single-topic sessions, ceasing to work, preening, or interrupting with requests such as "want cork." Testing would thus be impeded and test scores might decline "for reasons unrelated to competence" (Pepperberg, 2002, p. 56). Similarly, in an experiment to train a female killer whale to label visual objects with acoustic symbols, she would refuse to cooperate – and food made no difference – when her motivation decreased (Dudok van Heel, Kamminga, & van der Toorn, 1982).

Since prior to this set of experiments Qila had rarely been asked to repeat her vocal responses more than two or three consecutive times, the number of repeated trials in our experiments might have caused her to lose motivation to perform well. We examined this possibility by asking Qila to perform repeatedly a task she knew well: producing a scream or a pulse train in response to two different hand signals. She achieved a high level of success during these calibration trials, suggesting that her difficulty during the playback experiments was not a result of lack of motivation arising from the recurrent nature of the task. It could be argued, of course, that Qila simply disliked the vocal matching task compared to the calibration trials. However, she had ample opportunity to leave the training station. Had she disliked or become bored with the task, she could have simply swum away and refused to cooperate, but she did not. She stayed with us and was quite attentive throughout most sessions, even while performing poorly.

Learning to Learn

There is evidence that animals use prior learning experiences to facilitate subsequent learning—the *learning-to-learn* phenomenon, or *deutero-learning*. Bateson (1942, reprinted in Bateson, 1972, pp. 159-176) distinguished two types of learning: simple learning, denominated *proto-learning*, the adaptation of behavior in response to reinforcement contingencies, and *gestalt learning*, defined as deutero-learning, which refers to changes in the rate of proto-learning as a result of "insight" into the contexts of problem solving.

Schusterman (2008) described the process of learning-to-learn in one of his earlier experiments with a California sea lion. The author was testing whether the sea lion could transfer its knowledge of when to vocalize

and when to remain silent to many pairs of visual stimuli of different shapes and sizes. Every 160 trials a new problem was presented, for a total of 45 such problems. Interestingly, during the first five problems the subject averaged 59% correct responses, but in the following 10 problems correct responses jumped to 75%, reaching 88% correct responses on the last five problems. In other words, the sea lion solved each problem more rapidly than the earlier one (but note that 6400 trials were required for his best performance; many more than the number presented to Qila in this study). Similarly, the "creative porpoise" (Pryor, Haag, & Oreilly, 1969), a classic experiment in training for novelty and creativity, illustrated deutero-learning. A rough-toothed dolphin (*Steno bredanensis*) was rewarded only when it showed a new behavior. This took many sessions and seemed difficult for the animal until it understood the principle, after which new tricks came easily. Similarly, after years of training, at the time of writing the first draft of this paper (June 2016) both Qila and her mother Aurora had learned the concept of (and the hand signal for) "make a new vocalization," and readily offered varied vocalizations in response to this request. A good example of this phenomenon in humans is foreign language learning: a person learning a new language not only learns the language but also learns how to learn a language, becoming metalinguistically aware and thus acquiring subsequent languages more easily (Klein, 1995).

The concept of deutero-learning might be key in helping us understand Qila's performance. Qila was naive to research training. This was the first time that she was exposed to this kind of learning and had previously only been trained to perform a series of fairly simple behaviors for public shows. Previous experience participating in inherently more complex scientific research might have provided the necessary conceptual scaffolding to perform better in this experiment. Furthermore, in light of the enormous number of trials that some species require in order to make the cognitive switch that will suddenly accelerate their learning and render increasingly correct responses (e.g., sea lion experiment above), we probably did not run too many trials for Qila, but not enough. The fact that by the last two sessions of the transfer experiments she performed equally well matching both call types suggests that she might have improved with more sessions - unless this was simply an ephemeral, random peak (given the oscillation between sessions). Another indication that Qila may have had to "learn how to learn" may be her better performance with pulse trains, since during training, the scream vocalization was taught first, and only after she was able to match screams was the pulse train introduced.

At Kewalo Basin Marine Mammal Laboratory, six dolphins participated, for over 30 years, in a prodigious amount of research that produced ground-breaking discoveries on vocal and motor mimicry, language comprehension, memory, cross modal matching, sensory abilities, conceptual processes, and self awareness, amongst others (e.g., Beach & Herman, 1972; Herman, Pack, & Hoffman-Kuhnt, 1998; Herman, Richards, & Wolz, 1984; Herman, Uyeyama, & Pack, 2008; Pack & Herman, 1995, 2004; Ralston & Herman, 1995; Shyan & Herman, 1987; Thompson & Herman, 1977). These animals were embedded in an environment in which this kind of "education" was the norm; their entire lives were about learning in this manner. Animals consistently trained or taught in this manner simply have more experience interpreting the world according to the scientists' demands. In stark contrast to the beluga whale in Murayama et al.'s (2014) beluga vocal imitation study, who had previously undergone numerous behavioral and cognitive experiments, Qila had no such prior experience.

Experimental Limitations

Size of the training set. Stansbury et al. (2015) argued convincingly that small training sets impact a subject's ability to generalize to novel stimuli. They re-created Shapiro et al.'s (2004) earlier categorical matching study on two grey seal pups, which were trained successfully to respond to playbacks of moans and

growls with vocalizations of the same category but were unable to generalize this response. In contrast, Stansbury and colleagues showed that a grey seal pup trained with the same two stimulus classes (moans and growls) but with a larger training set (212 exemplars, compared to nine in Shapiro et al.'s study) could indeed discriminate between novel calls of the two classes. The fact that several successful studies on category formation have used large training sets (e.g., bonobos, *Pan paniscus*: Pedersen, 2012; African grey parrot, *Psittacus erithacus*: Pepperberg & Brezinsky, 1991; domestic dogs, *Canis familiaris*: Range, Aust, Steurer, & Huber, 2008) may indicate that the increased number of initial stimuli may allow the subject to more easily perceive and corroborate the perceptual coherence of the different categories (Wasserman, Brooks, & McMurray, 2015). In our study, Qila had difficulty with both the trained set and the transfer set, indicating that she may have been attempting to learn the response to each specific call in the first place, rather than generalizing the calls into categories to accomplish the task.

Number of elements in each call exemplar. The fact that we chose to play sequences of calls (two pulse trains or four screams) rather than a single call exemplar during each playback meant that Qila could have learned to discriminate simply by counting the numbers of sounds heard. Although the results do not suggest this was the case given Qila's poor performance, this experimental flaw should be considered and improved upon in future experiments.

Order of presentation of the playback stimuli. Although the presentation of the playback stimuli was randomized during both the trained-set and transfer-set trials, this was not the case during the training phase prior to the categorical matching experiment. The presentation of playback stimuli was not randomized then, but was dictated by the trainers. Since Qila was given a second presentation of stimuli to which she had given an incorrect response (in order to give Qila the chance to match the vocalization that she got wrong, to receive a reward and make the training as positive as possible), she could have learned that if she gave an incorrect response the first time, she could get a reward by simply switching to the other vocalization type. The fact that she sometimes offered both vocalizations in response to screams raises the possibility that she may have been using a different set of rules in making her responses rather than trying to correctly match the call type category.

Procedural limitations. It is possible that Qila's poor performance stems from procedural limitations rather than from a true difficulty performing the task. The training method - operant conditioning with food reinforcement - might not be conducive to stimulating vocal learning in these animals. Several authors have proposed that food may not be the best type of reinforcer to control vocal output when vocalizations in nature are largely under the control of social reinforcers (Adret, 1993; Manabe et al., 1997; Pierce, 1985). Belugas have developed an elaborate system of vocal signals to mediate their social interactions. This begs the consideration that perhaps social reinforcers should be expected to influence their vocal behavior to a larger extent than food.

The idea that animal vocalizations are highly susceptible to training protocols that emphasize interaction and social rewards has gathered a great deal of evidence from studies on grey parrots (*Psittacus erithacus*) (e.g., Pepperberg, 1985, 1993, 1999; Pepperberg, Gardiner, & Luttrell, 1999; Pepperberg & Neapolitan, 1988; Todt, 1975). Pepperberg (1985, 1993) used human social modeling theory as a framework to guide her avian vocal learning studies, and showed that exposing parrots to the presence of interacting human "tutors" clearly accelerates the learning and appropriate usage of new vocal labels and concepts.

Pepperberg's experimental paradigm has parallels in cetacean research. Reiss and McCowan (1993) manipulated the acoustic environment of subjects sufficiently to observe spontaneous (untrained) mimicry

without altering their social environment. They designed a large underwater keyboard that stimulated the dolphin's natural curiosity, so that the animals spontaneously pressed visual forms on the keyboard, which were linked to different synthetic whistles. Each whistle was in turn linked to a specific object (e.g., ball, ring) or activity (e.g., rub). The rapidity with which the dolphins mimicked the model sounds and used them in appropriate contexts contrasted greatly with the results reported by previous studies (Richards, Woltz, & Herman, 1984; Sigurdson, 1993), which used traditional food reinforcement methods and required many trials to train initial vocal mimicry. This suggests that a social and interactive environment, which also allows for spontaneity, whereby the animal may initiate the interaction, might be more conducive to the learning of new signals, categorical matching, and mimicry than training protocols divorced of social factors.

A more recent study by Hooper, Reiss, Carter, and McCowan (2006) corroborated some of McCowan and Reiss's earlier results. They exposed two infant male bottlenose dolphins and their mothers to six novel computer-generated whistles that were either unpaired or paired with toy objects. The results showed that the dolphins imitated those whistles that were paired with objects after fewer exposures, more frequently, and producing higher quality imitations than those that were unpaired, indicating that context matters.

Herman (2002), recounted the three decades of prolific studies on dolphin cognitive, sensory and linguistic competencies, noting that *teaching* models were often more efficient and effective than *training* models based on traditional instrumental conditioning techniques. He pointed out that these teaching models employed techniques that are commonly used with young children, such as showing and demonstrating.

Qila's poor performance in the categorical matching experiment might also be explained if this species' perceptual capacity of sounds broadcasted in air is reduced. Auditory systems of cetaceans are highly specialized for underwater hearing, whereas their hearing capacity in air is not well documented (Liebschner, Hanke, Miersch, Dehnhardt, & Sauerland, 2005). Cetacean species for which hearing sensitivity to airborne sounds was investigated (e.g., bottlenose dolphins: Babushina, 1979; harbour porpoises, *Phocoena phocoena*: Kastelein, Nieuwstraten, Staal, Van Ligtenberg, & Versteegh, 1997; tucuxis, *Sotalia fluviatilis guianensis*: Liebschner et al., 2005) showed better hearing underwater sensitivity compared to aerial, for all frequencies tested, when using sound intensity. However, when aerial and underwater sound detection thresholds were compared via sound pressure instead of intensity, the results were less clear-cut, with two of the species (*Tursiops* and *Sotalia*) showing more sensitivity to airborne sounds at particular frequencies (summarized in Liebschner et al., 2005).

Belugas have a wide auditory bandwidth, with reported hearing ranges falling within 1-150 kHz (Awbrey, Thomas, & Kastelein, 1988; Finneran et al., 2005; Klishin, Popov, & Supin, 2000; White, Norris, Ljungblad, Barton, & di Sciara, 1978) and a frequency bandwidth of best hearing from 22.5 to 80 kHz according to the most recent Auditory Evoked Potential study of seven wild temporarily restrained belugas (Castellote et al., 2014). However, the extent of their hearing sensitivity to airborne sounds is unknown, so the question of whether her performance was affected by the fact that the playbacks were broadcasted in-air must, for the moment, remain unanswered.

The problem inherent in this sort of experiment is that a high above-chance response provides an unequivocal answer that a particular capability is present, but a poor or near-chance performance does not provide an equally unequivocal answer that the capability is absent, because we are unsure about the extent to which various procedural factors limited success. McGreevy and Boakes (2007) question the tacit belief that an individual animal that is very trainable in a particular task has a better understanding of this task than another individual that was difficult to train. Instead, they argue, variability in individuals' responses to training

probably reflects various factors, including the trainer's skills. A combination of the issues explored here might explain Qila's difficulty.

Experimental limitations notwithstanding, the idea that Qila had a true difficulty perceiving categorically a signal that lacks a function in the natural repertoire of belugas is worth exploring, as is the notion of differences in perceptual boundaries between cetaceans and humans affecting a subject's ability to discriminate between categories of sound.

Acknowledgments

References

Adret, P. (1993). Vocal learning induced with operant techniques: An overview. *Netherlands Journal of Zoology, 43*, 125–142.

Au, W. W. L., & Hastings, M. C. (2009). *Principles of marine bioacoustics*. New York, NY: Springer. doi:10.1007/978-0-387-78365-9

Awbrey, F. T., Thomas, J. A., & Kastelein, R. A. (1988). Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*. *Journal of the Acoustical Society of America, 84*, 2273–2275.

Bates, D., & Sarkar, D. (2006*). lme4: Linear mixed effects models using S4 classes*. Retrieved from http://CRAN.R-project.org, R package version 0.99875-8.

Babushina, Y. S. (1979). Localization by the dolphin of the source of tonal and pulse signals in water and air. *Vestnik Leningradskogo Universiteta Biologiva*, *3*, 119–121.

Bateson, G. (1972). Steps to an ecology of mind. San Francisco, CA: Chandler Publishing Co.

Baugh, A. T, Akre, K. L., & Ryan, M. J. (2008). Categorical perception of a natural, multivariate signal: Mating call recognition in túngara frogs. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 8985–8988. doi:10.1073/pnas.0802201105

Beach, F. A. I., & Herman, L. M. (1972). Preliminary studies of auditory problem solving and intertask transfer by the bottlenose dolphin. *The Psychological Record, 22,* 49–62.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H., & White, J. S. (2008). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution, 24,* 127–135. doi:10.1016/j.tree.2008.10.008

Bradbury, J. W. (2003). Vocal communication in wild parrots. In F.B.M. de Waal, & P. Tyack (Eds.), *Animal Social Complexity: Intelligence, Culture, and Individualized Societies (*pp. 227–233). Cambridge, MA: Harvard University Press.

Caldwell, M. C., & Caldwell, D. K. (1967). Intra-specific transfer of information via pulsed sound in captive odontocete cetaceans. In R. G. Busnel (Ed.). *Animal Sonar Systems* (pp. 879–836). Jouy-en-Josas, France : Laboratoire de Physiologie Acoustique.

Castellote, M., Mooney, A. T., Quakenbush, L., Hobbs, R., Goertz, C., & Gaglione, E. (2014). Baseline hearing abilities and variability in wild beluga whales (*Delphinapterus leucas*). *Journal of Experimental Biology*, *217*, 1682–91. doi:10.1242/jeb.093252

Cheney, D. L., & Seyfarth, R. M. (1982). How vervet monkeys perceive their grunts - Field playback experiments. *Animal Behaviour, 30*, 739–751. doi:10.1016/S0003-3472(82)80146-2.

Davidson, L., Shaw, J., & Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *Journal of the Acoustical Society of America, 122,* 3697–3709. doi:10.1121/1.2801548

Dudok van Heel, W. H., Kamminga, C., & van der Toorn, J. D. (1982). An experiment in two-way communication in *Orcinus orca*. *Aquatic Mammals*, *9,* 69–82.

Eaton, R. L. (1979). A beluga whale imitates human speech. *Carnivore*, *2,* 22–23.

Egnor, S. E. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences, 27,* 649–654. doi:10.1016/j.tins.2004.08.009

Ehret, G., & Haack, B. (1981). Categorical perception of mouse pup ultrasound by lactating females. *Naturwissenschaften, 68,* 208–209.

Finneran, J. J., Carder, D. A., Dear, R., Belting, T., McBain, J., Dalton, L. & S. H. Ridgway. (2005). Pure tone audiograms and possible aminoglycoside-induced hearing loss in belugas (*Delphinapterus leucas*). *Journal of the Acoustical Society of America*, *117,* 3936–3943. doi: 10.1121/1.1893354

Fischer, J. (1998). Barbary macaques categorize shrill barks into two call types. *Animal Behaviour, 55,* 799–807. doi:10.1006/anbe.1997.0663

Ginsburg, N. (1963). Conditioned talking in the mynah bird. *Journal of Comparative and Physiological Psychology, 56*, 1061–1063.

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signaling in the recruitment of agonistic aid. *Animal Behaviour, 32,* 182–193. doi:10.1016/S0003-3472(84)80336-X

Hall, M. L. (2004). A review of hypotheses for the functions of avian duetting. *Behavioral Ecology and Sociobiology*, *55*, 415–430. doi:10.1007/s00265-003-0741-x

Hauser, M. D. (1989). Ontogenetic changes in the comprehension and production of vervet Monkey (*Cercopithecus aethiops*) vocalizations. *Journal of Comparative Psychology*, *103,* 149–158. doi:10.1037/0735-7036.103.2.149

Hauser, M. D. (1998). Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys. *Animal Behaviour, 55,* 1647–1658. doi:10.1006/anbe.1997.0712

Herman, L. M. (2002). Exploring the cognitive world of the bottlenose dolphin. In M. Bekoff, C. Allen, & G. Burhardt (Eds.), The *cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 275–283). Cambridge, MA: MIT Press.

Herman, L. M., Pack, A., & Hoffman-Kuhnt, M. (1998). Seeing through sound: Dolphins (*Tursiops truncatus*) perceive the spatial structure of objects through echolocation. *Journal of Comparative Psychology*, *112*, 392–305. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9770316

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition, 16,* 129–219. doi:10.1016/0010-0277(84)90003-9

Herman, L. M., Uyeyama, R. K., & Pack, A. A. (2008). Bottlenose dolphins understand relationships between concepts. *Behavioral and Brain Sciences, 31*, 139–140. doi:10.1017/S0140525X08003658

Hooper, S., Reiss, D., Melissa, C. & McCowan, B. (2006). Importance of contextual saliency on vocal imitation by bottlenose dolphins. *International Journal of Comparative Psychology*, *19*, 116–128. Retrieved from http://escholarship.org/uc/item/1nw8p7jt

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science*, *289,* 1355–1357. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10958783

Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behaviour, 26,* 59–99.

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour, 60,* 1–11. doi:10.1006/anbe.2000.1410

Kastelein, R. A., Nieuwstraten, S. H., Staal, C., van Ligtenberg, C. L., & Versteegh, D. (1997). Low-frequency aerial hearing of a harbour porpoise (*Phocoena phocoena*). In A. J. Read, P. R. Wiepkema, & P.E. Nachtigall, (Eds.). *The Biology of the Harbour Porpoise (*pp. 295–312). Woerden, The Netherlands: De Spil Publishers.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 11381–11386. doi: 10.1073/pnas.0604821103

Klein, E. C. (1995). Second versus third language acquisition: is there a difference? *Language Learning, 45*, 419–465.

Klishin, V. O., Popov, V. V., & Supin A. Y. (2000). Hearing capabilities of a beluga whale, *Delphinapterus leucas. Aquatic Mammals, 26,* 212–228.

Le Prell, C. G., Hauser, M. D., & Moody, D. B. (2002). Discrete or graded variation within rhesus monkey screams*? Psychophysical experiments on classification. Animal Behaviour*, *63* ,47–62. doi:10.1006/anbe.2001.1888

Liebschner, A., Hanke, W., Miersch, L., Dehnhardt, G., & Sauerland, M. (2005). Sensitivity of a tucuxi (*Sotalia fluviatilis guianensis*) to airborne sound. *Journal of the Acoustical Society of America, 117,* 436–441. doi:10.1121/1.1828851

Manabe, K., Kawashima, T., & Staddon, J. E. R. (1995). Differential vocalization in budgerigars: Towards an experimental analysis of naming. *Journal of the Experimental Analysis of Behavior*, *63,* 111–126.

Manabe, K., Staddon, J. E. R., & Cleaveland, J. M. (1997). Control of the vocal repertoire by reward in budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology, 111,* 50–62. doi:10.1037/0735-7036.111.1.50

McComb, K., Reby, D., Baker, L., Moss, C., & Sayialel, S. (2003). Long-distance communication of acoustic cues to social identity in African elephants. *Animal Behaviour, 65,* 317–329. doi:10.1006/anbe.2003.2047

McGreevy, P. D., & Boakes, R. A. (2007). *Carrots and sticks: Principles of animal training.* New York, NY Cambridge University Press.

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., & Solow, A. R. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. *Animal Behaviour*, *67,* 1099–1107. doi:10.1016/j.anbehav.2003.06.017

Murayama, T., Iijima, S., Katsumata, H., & Aai, K. (2014). Vocal Imitation of Human Speech, Synthetic Sounds and Beluga Sounds, by a Beluga (*Delphinapterus leucas*). *International Journal of Comparative Psychology, 27,* 369–384. Retrieved from http://escholarship.org/uc/item/51v1z12b

Murray, S. O., Mercado, E., & Roitblat, H. L. (1998). Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America, 104,* 1679–1688. Retrieved from http://dx.doi.org/10.1121/1.424380

Nelson, D. A., & Marler, P. (1989). Categorical perception of a natural stimulus continuum: Birdsong. *Science, 244,* 976–978.

Owren, M. J. (1990). Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*):1. Natural calls. *Journal of Comparative Psychology, 104,* 20–28. Retrieved from http://dx.doi.org/10.1037/0735-7036.104.1.20

Pack, A., & Herman, L. M. (1995). Sensory integration in the bottlenose dolphin: Immediate recognition of complex shapes across the senses of echolocation and vision. *Journal of the Acoustical Society of America, 98*, 722–733. Retrieved from http://dx.doi.org/10.1121/1.413566

Pack, A. A., & Herman, L. M. (2004). Bottlenosed dolphins (*Tursiops truncatus*) comprehend the referent of both static and dynamic human gazing and pointing in an object-choice task. *Journal of Comparative Psychology, 118,* 160–171. doi:10.1037/0737036.118.2.160

Pedersen, J. (2012). *The symbolic mind: Apes, symbols, and the evolution of language* (Doctoral dissertation). Iowa State University, Ames, Iowa.

Pepperberg, I. M. (1985). Social modeling theory: A possible framework for understanding avian vocal learning. *Auk*, *102,* 854–864.

Pepperberg, I. M. (1993). A review of the effects of social interactions on vocal learning in African grey parrots (*Psittacus erithacus*). *Netherlands Journal of Zoology, 43,* 104–124.

Pepperberg, I. M. (1999). The *Alex studies: Cognitive and communicative abilities of grey parrots.* London, UK: Harvard University Press.

Pepperberg, I. M. (2002). In search of King Solomon's ring: Cognitive and communicative studies of grey parrots (*Psittacus erithacus*). *Brain Behavior and Evolution, 59,* 54–67. doi:10.1159/000063733

Pepperberg, I. M., & Brezinsky, M. V. (1991). Acquisition of a relative class concept by an African gray parrot (*Psittacus erithacus*): Discriminations based on relative size. *Journal of Comparative Psychology, 105,* 286–294. doi:10.1037/0735-7036.105.3.286

Pepperberg, I. M., & Neapolitan, D. M. (1988). Second language acquisition: A framework for studying the importance of input and interaction in exceptional song acquisition. *Ethology, 77,* 150–168. doi:10.1111/j.1439-0310.1988.tb00199.x

Pepperberg, I. M., Gardiner, L. I., & Luttrell, L. J. (1999). Limited contextual vocal learning in the grey parrot (*Psittacus erithacus*): The effect of interactive co-viewers on videotaped instruction. *Journal of Comparative Psychology, 113,* 735–736. doi:10.1037/0735-7036.113.2.158

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention Perception & Psychophysics, 71,* 1385–1398.

Pierce, J. D. J. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates, 26,* 202–213. doi:10.1007/BF02382019

Pryor, K. W., Haag, R., & Oreilly, J. (1969). The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of Behavior, 12,* 653–661. doi: 10.1901/jeab.1969.12–653

Ralston, J. V., & Herman, L. M. (1995) Perception and generalization of frequency contours by a bottlenose dolphin (*Tursiops truncatus*). *Journal of Comparative Psychology, 109,* 268–277. doi:10.1037/0735-7036.109.3.268

Range, F., Aust, U., Steurer, M., & Huber, L. (2008). Visual categorization of natural stimuli by domestic dogs. *Animal Cognition, 11,* 339–347. doi:10.1007/s10071-007-0123-2

Reiss, D., & McCowan, B. (1993). Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning. *Journal of Comparative Psychology, 107,* 301–312. doi: 10.1037/0735-7036.107.3.301

Richards, D. G., Woltz, J. P., & Herman, L. M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin (*Tursiops truncatus*). *Journal of Comparative Psychology, 98,* 10–28. doi:10.1037/0735-7036.98.1.10

Ridgway, S. H., Carder, D. A., & Jeffries, M. M. (1985). Another "talking" male white whale. *Abstracts from the Sixth Biennial Conference on the Biology of Marine Mammals*, Vancouver, B. C., Canada, Nov.22-26, p. 67

Sayigh, L.S. (1992). *Development and function of signature whistles of free-ranging bottlenose dolphins, Tursiops truncatus* (Doctoral dissertation). Woods Hole Oceanographic Institution and Massachusetts Institute of Technology, Woods Hole.

Schulz, T. M., Whitehead, H., Gero, S, & Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: Insights into communication function. *Animal Behaviour*, *76*, 1977–1988. doi:10.1016/j.anbehav.2008.07.032

Schusterman, R. J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In D. K. Oller & U. Gribel (Eds.), *The evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication* (pp. 41–70). Cambridge, MA: MIT Press.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls - Semantic communication in a free-ranging primate. *Animal Behaviour, 28,* 1070–1094. doi:10.1016/S0003-3472(80)80097-2

Shapiro, A. D., Slater, P. J. B., & Janik, V. M. (2004). Call usage learning in gray seals (*Halichoerus grypus*). *Journal of Comparative Psychology, 118*, 447–454. doi:10.1037/0735-7036.118.4.447

Shyan, M. R., & Herman, L. M. (1987). Determinants of recognition of gestural signs in an artificial language by Atlantic bottle-nosed dolphins (*Tursiops truncatus*) and humans (*Homo sapiens*). *Journal of Comparative Psychology, 101,* 112–125. doi:10.1037/0735-7036.101.2.112

Sigurdson, J. (1993). Whistles as a communication medium. In H. L. Roitblat, L. M. Herman, & P. E. Nachtigall (Eds.), *Language and Communication: Comparative Perspectives* (pp. 153–173). Hillsdale, NJ: Erlbaum Associates.

Sinnott, J. M. (1987). Modes of perceiving and processing information in birdsong (*Agelaius phoeniceus*, *Molothrus ater*, and *Homo sapiens*). *Journal of Comparative Psychology, 101,* 355–366.

Snowdon, C. T., & Pola, Y. V. (1978). Interspecific and intraspecific responses to synthesized Pygmy marmoset vocalizations. *Animal Behaviour, 26,* 192–206. doi:10.1016/0003-2472(78)90019-2

Stansbury, A. L., de Freitas, M., Wu, G. M., & Janik, V. M. (2015). Can a gray seal (*Halichoerus grypus*) generalize call classes? *Journal of Comparative Psychology, 129*, 412–420. doi:10.1037/a0039756

Sugiura, H., & Masataka, N. (1995). Temporal and acoustic flexibility in vocal exchanges of coo calls in Japanese macaques (*Macaca fuscata*). In U. Jürgens, J. Newman, & E. Zimmermann (Eds.), *Current topics in primate vocal communication* (pp. 121–140). New York, NY: Plenum Press.

Teixidor, P., & Byrne, R. W. (1999). The 'whinny' of spider monkeys: Individual recognition before situational meaning. *Behaviour, 136,* 279–308. Retrieved from: http://www.jstor.org/stable/4535610

Thompson, R. K. R., & Herman, L. M. (1977). Memory for lists of sounds by the bottle-nosed dolphin: Convergence of memory processes with humans? *Science, 195,* 501–503.

Todt, D. (1975). Social learning of vocal patterns and modes of their applications in grey parrots (*Psittacus erithacus*). *Ethology, 39*, 178–188. doi: 10.1111/j.1439-0310.1975.tb00907.x

Turl, C. W., & Penner, R. H. (1989). Differences in echolocation click patterns of the beluga (*Delphinapterus leucas)* and the bottlenose dolphin (*Tursiops truncatus). Journal of the Acoustical Society of America, 86,* 497–502.

Turl, C. W., Penner, R. H., & Au, W. W. L. (1987). Comparison of target detection capabilities of the beluga and bottlenose dolphin. *Journal of the Acoustical Society of America, 85*, 1487–1491.

Turl, C. W., Skaar, D. J., & Au,W. W. L. (1991). The echolocation ability of the beluga (*Delphinapterus leucas*) to detect targets in clutter. *Journal of the Acoustical Society of America*, *89*, 896–901.

Vergara, V. & Barrett-Lennard, L. G. (2008). Vocal development in a beluga calf (*D. leucas*). *Aquatic Mammals, 34,* 123–143. doi:10.1578/AM.34.1.2008.123

Vergara, V., Michaud, R., & Barrett-Lennard, L. G. (2010). What can captive whales tell us about their wild counterparts? Identification, usage and ontogeny of contact calls in belugas (*Delphinapterus leucas*). *International Journal of Comparative Psychology*, *23,* 278–309. Retrieved from: http://escholarship.org/uc/item/4gt03961

Waser, P. M. (1977). Individual recognition, intragroup cohesion and intergroup spacing: Evidence from sound playback to forest monkeys. *Behaviour*, *60,* 28–74. Retrieved from: http://www.jstor.org/stable/4533792

Wasserman, E. A., Brooks, D. I., & McMurray, B. (2015). Pigeons acquire mulitple categories in parallel via associative learning: A parallel to human word learning? *Cognition, 136*, 99–122. doi:10.1016/j.cognition.2014.11.020.

Weary, D. M. (1989). Categorical perception of bird song - How do great tits (*Parus major*) perceive temporal variation in their song*? Journal of Comparative Psychology, 103,* 320–325. doi:10.1037/0735-7036.103.4.320

White, M. J., Norris, J. C., Ljungbland, D. K., Barton, K., & di Sciara, G. N. (1978). Auditory thresholds of two beluga whales (*Delphinapterus leucas*) (Hubbs/Sea World Research Institute Technical Report H/SWRI 78-109). San Diego, CA: Hubbs Marine Research Institute.