

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Can a gray seal (*Halichoerus grypus*) generalize call classes?

Amanda L. Stansbury, Mafalda de Freitas, Gi-Mick Wu, Vincent M. Janik

Sea Mammal Research Unit, University of St. Andrews, Fife KY16 8LB, UK

Corresponding author: Amanda Stansbury, as252@st-andrews.ac.uk, Sea Mammal Research
Unit, University of St. Andrews, Fife KY16 8LB, UK.

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Abstract

Past research found that grey seals (*Halichoerus grypus*) were capable of classifying vocal signals by call type using a trained set, but were unable to generalize to novel exemplars (Shapiro, Slater, & Janik, 2004). Given the importance of auditory categorization in communication, it would be surprising if the animals were unable to generalize acoustically similar calls into classes. Here, we trained a juvenile grey seal to discriminate novel calls into two classes, ‘growls’ and ‘moans’, by vocally matching call types (i.e., the seal ‘moaned’ when played a ‘moan’ and ‘growled’ when played a ‘growl’). Our method differed from the previous study as we trained the animal using a comparatively large set of exemplars with standardized durations, consisting of both the seal’s own calls and those of two other seals. The seal successfully discriminated ‘growls’ and ‘moans’ for both her own (94% correct choices) and other seal’s (87% correct choices) calls. A generalized linear model (GLM) found that the seal’s performance significantly improved across test sessions, and accuracy was higher during the first presentation of a sound from her own repertoire but decreased after multiple exposures. This pattern was not found for calls from unknown seals. Factor analysis for mixed data (FAMD) identified acoustic parameters that could be used to discriminate between call types and individuals. ‘Growls’ and ‘moans’ differed in noise, duration and frequency parameters, whereas individuals differed only in frequency. These data suggest that the seal could have gained information about both call type and caller identity using frequency cues.

Keywords: Classification, generalization, discrimination, gray seal, call types

A SEAL GENERALIZES CALL CLASSES

51 Can a gray seal (*Halichoerus grypus*) generalize call classes?

52 A fundamental ability in animal cognition is the capacity for classification and
53 generalization of stimuli. Classifying objects reduces large numbers of stimuli into a small
54 group of categories, allowing animals to cope with stimulus variability. Generalization
55 places new stimuli into pre-existing categories, speeding up signal processing and response
56 time.

57 Classification and generalization of stimuli has been well investigated in the visual
58 domain for several species, such as bottlenose dolphins (e.g., Mercado, Killebrew, Pack,
59 Mácha, & Herman, 2000), domestic dogs (Range, Aust, Steurer, & Huber, 2008), pigeons
60 (e.g., Wasserman, Zentall, & Cook, 2006), parrots (e.g., Pepperberg, 1996, 1999) and
61 nonhuman primates (e.g., Neiwirth & Wright, 1994; Schrier & Brady, 1987). Many species
62 are similarly capable of auditory categorization. For example animals can use sound
63 categorization to classify objects, particularly echolocating species (e.g., Helverson, 2004;
64 Roitblat, Penner, & Nachtigall, 1990). Past research has mainly concentrated on the
65 classification of biological sounds, such as bird song (e.g., Beecher, Campbell, & Burt, 1994)
66 and predator-specific alarm calls (e.g., Cäsar, Zuberbühler, Young, & Byrne, 2013; Greene &
67 Meagher, 1998; Seyfarth, Cheney, & Marler, 1980). Some species categorize conspecific
68 call types by responding to another individual's calls with a matching call type, referred to as
69 antiphonal calling (e.g., Ghazanfar, Flombaum, Miller, & Hauser, 2001; Soltis, Leong, &
70 Savage, 2005). Perhaps most importantly, auditory categorization helps animals to deal with
71 individual variation in communication calls, where classes of signals have different meaning.
72 Previous studies have used playback tests and discrimination training to find out how
73 individuals categorize sounds, providing insight into how animals judge conspecific (e.g.,
74 Candiotti, Zuberbühler, & Lemasson, 2013; Lind, Dabelsteen, & McGregor, 1996; Maros et
75 al., 2008; Weary & Krebs, 1992) and/or heterospecific sounds, including human speech and

A SEAL GENERALIZES CALL CLASSES

76 music (e.g., Bloomfield, Sturdy, Phillmore, & Weisman, 2003; Brooks & Cook, 2010;
77 Hoeschele, Cook, Guillette, Brooks, & Sturdy, 2012; Kuhl, 1981; Pepperberg, 1999; Porter &
78 Neuringer, 1984). Several species are able to discriminate between hundreds of human
79 speech sounds and generalize words produced by different speakers (e.g., dogs: Kaminski,
80 Call, & Fischer, 2004; parrots: Pepperberg, 1999; primates: Savage-Rumbaugh et al., 1993).

81 Given the importance of auditory categorization in establishing a common
82 communication code, it would be surprising if animals were unable to generalize calls that
83 are acoustically similar. Nevertheless, such a result has been reported for grey seals. Shapiro
84 et al. (2004) found that two juvenile grey seals, “Kylie” and “Oscar”, were capable of
85 discriminating between two call types: ‘growls’ and ‘moans’. These call types are easily
86 distinguishable by human observers. ‘Growls’ are broadband, noisy calls with a bandwidth
87 upper limit extending to frequencies up to 20 kHz, whereas ‘moans’ are tonal, periodic calls
88 with a harmonic structure and a bandwidth upper limit rarely exceeding 5 kHz (Shapiro et al.,
89 2004; Figure 1). Kylie and Oscar were trained to vocalize upon the presentation of a sound:
90 recordings of the seals’ own ‘growls’ and ‘moans’. The seals were trained with a set of nine
91 exemplars (five ‘growls’ and four ‘moans’ for Kylie, vice-versa for Oscar), and were
92 reinforced for responding with a matching call type (i.e., responded with a ‘moan’ when
93 played a ‘moan’ and a ‘growl’ for a ‘growl’). Both seals successfully learned the task after
94 several hundred trials (~800 trials for Oscar and ~1,650 for Kylie). However, neither seal
95 accurately responded when presented with novel exemplars (new sets of their own ‘growls’
96 and ‘moans’). One seal responded variably to the new calls, whereas the other consistently
97 responded with a ‘growl’. Shapiro et al. (2004) concluded that although their seals were able
98 to discriminate between call types, they were unable to generalize classification to novel call
99 exemplars.

A SEAL GENERALIZES CALL CLASSES

100 There are several possible explanations for why the seals were unable to generalize
101 novel exemplars, such as the training set size and parameter differences between training and
102 testing stimuli. Training set size has been previously shown to impact generalization; for
103 example, larger training sets allow for faster category formation and improved accuracy (e.g.,
104 Katz & Wright, 2006; Schrier & Brady 1987; Wasserman & Bhatt 1992; Wright, Cook,
105 Rivera, Sands & Delius, 1988; Wright, Rivera, Katz, & Bachevalier, 2003). Shapiro et al.
106 (2004) used a small, fixed set of stimuli (nine exemplars) to train classification before testing
107 generalization with new stimuli. Although such small training sets have been successfully
108 used for discrimination tasks in other species (e.g., Brooks & Cook, 2010; Watanabe, 2001),
109 it is possible that the seals simply learned through trial and error to respond with the
110 appropriate call type for each separate stimulus, as opposed to generalizing the calls into
111 categories. This is consistent with results from other auditory classification studies which
112 have been similarly unsuccessful at training call type categorization of novel stimuli in rhesus
113 monkeys (*Macaca mulatta*: Le Prell, Hauser, & Moody, 2002) and a beluga whale
114 (*Delphinapterus leucas*: Vergara, 2011). Both studies utilized small sets of training stimuli
115 (eight exemplars of each type for the rhesus monkeys, three of each for the beluga) before
116 testing generalization to novel sounds. Other classification tasks have utilized large, variable
117 sets of stimuli. In a notable example, an African grey parrot (*Psittacus erithacus*), ‘Alex’,
118 was successfully trained to classify objects in several ways including colour, size and shape
119 using dozens of training exemplars (Pepperberg, 1999). For example, during a size
120 discrimination task Alex was required to identify which object in a pair was either larger or
121 smaller (Pepperberg & Brezinsky, 1991). Initially, the training set consisted of eighteen
122 exemplars but was then expanded to include thirty-two different types of objects, with
123 exemplars of each type varying in size and colour. By using a larger exemplar set, it is

A SEAL GENERALIZES CALL CLASSES

124 unlikely that Alex learned specific responses to specific stimuli. Instead it is more likely he
125 generalized across exemplars to accomplish the task (Pepperberg & Brezinsky, 1991).

126 Another possible explanation for Shapiro et al.'s (2004) result is that the duration of
127 signals changed between the training set and novel stimuli. For example, during training
128 Oscar was played 'moans' with an average duration of 1.5 seconds (± 0.9) and 'growls' of 2.6
129 seconds (± 1.2). For the novel exemplars, the absolute duration of both call types changed
130 (novel 'moans' were averagely 0.4 seconds ± 0.1 in length, and 'growls' 0.7 seconds ± 0.1).
131 Although the relative duration remained consistent ('moans' were roughly half the duration
132 of 'growls' for both training and novel stimuli), if absolute duration was used as a cue to
133 discriminate between call types, this would prevent the seals from successfully performing
134 the task.

135 Additionally, Shapiro et al. (2004) only presented the seals with their own sounds.
136 Although this method does test the animal's ability to classify its own calls, the seal should
137 also be capable of generalizing novel exemplars from conspecifics. For auditory
138 categorization to be used for communication, animals must both be able classify the calls
139 they produce as well as the calls they perceive from others. Thus, generalization should be
140 tested for both the seal's own calls as well as for those of other individuals.

141 The present study tested a juvenile grey seal's ability to discriminate her own sounds
142 in addition to those produced by other, unknown juvenile grey seals. We report that the seal
143 was capable of generalizing two vocalization classes, 'growls' and 'moans', when trained
144 using a large, continually changing sets of stimuli with standardized durations.

145 **Methods**

146 **Subject**

147 A juvenile female grey seal, "Zola" (tag numbers 73254/5), born November 7, 2011
148 on the Isle of May (Firth of Forth, Scotland), was the subject of this study. After weaning,

A SEAL GENERALIZES CALL CLASSES

149 Zola was transported to our laboratory (St. Andrews, Scotland) on December 6th, 2011. She
150 was housed with four other juveniles in three enclosures including a large rectangular pool
151 (42m x 6m x 2.5m) and two circular pools (3m x 5m x 2m). The seals were fed a varied diet
152 of several fish species (mostly consisting of herring, *Clupea harengus*, and sprat, *Clupea*
153 *sprattus*). Training and testing occurred at the facility for twelve months, after which Zola
154 was released into the wild.

155 **Acoustic Recordings**

156 Acoustic recordings were obtained using a Sennheiser MKH 416 P48 directional
157 microphone (frequency response 40-20,000 Hz \pm 1 dB) and an Edirol FA-66 external sound
158 card (sampling rate 96 kHz, 24-bit) with a laptop computer. Weather permitting, sessions
159 were concurrently video recorded using a Sony HDR CX250E video camera. Stimuli were
160 played from the laptop using the FA-66 through an external Skytec 170.170 active speaker
161 (frequency response 32-22,000 Hz). Sounds were simultaneously played, recorded and
162 spectrographically monitored in real time using the program Audacity 1.3 (sampling rate 96
163 kHz, 24-bit, Audacity Team, 2012).

164 **Training and Testing Procedure**

165 Positive reinforcement behavioural training was conducted using the seal's normal
166 daily diet as a reward. Husbandry training began in January 2012 and focused on general
167 behaviours including exiting the water and stationing at the poolside for testing. Vocalization
168 training began in March 2012. Initially Zola was reinforced for any sound she produced
169 while in the same enclosure as the other seals. This was done because the seals vocalized
170 more frequently while in a group, and were relatively silent when isolated. Once Zola
171 regularly vocalized, she was separated from the others for individual sessions. At this time
172 Zola produced the two distinct call types, 'growls' and 'moans'. Each call type was paired
173 with a unique hand cue. The hand cues were then faded out to be replaced with the

A SEAL GENERALIZES CALL CLASSES

174 presentation of a sound (either a ‘growl’ or ‘moan’) and a light. The LED light was used to
175 indicate when the sound had finished and cue the seal’s response. After the hand cues were
176 replaced by sound, all training and testing sessions occurred with the experimenter waiting
177 quietly out of sight.

178 Simultaneous to training the vocal responses, the trainer reinforced Zola for staying
179 quiet when the vocal cues were not presented. This behavior was shaped until she remained
180 quietly at station between trials for up to fifteen minutes at a time. Although between most
181 trials this stationing period was less than a minute, in some cases the inter-trial interval was
182 extended to avoid an overlap with environmental noise. Training sessions were variable in
183 length depending on the seal’s performance; a set amount of food was used to reinforce the
184 seal, and sessions continued until all the fish was used (mean number of trials per session =
185 61, $SD \pm 26$, mean session duration = 37 minutes, $SD \pm 9$).

186 Throughout training and testing, each trial was initiated by the seal leaving the water
187 and stationing in front of the speaker. Once positioned, a sound (either a ‘growl’ or a
188 ‘moan’) was played. After the sound ended, an LED light illuminated and cued the seal’s
189 response. The animal was required to produce a vocalization of the same type. This response
190 was then evaluated using a visual judgement of a real-time spectrogram display comparing
191 the played stimuli to the seal’s response in Audacity. If correct (produced same call type as
192 the one played), the seal was reinforced with fish after it stopped vocalizing. If incorrect
193 (produced different call type than the one played), a time out of five to ten seconds was used
194 before the seal could initiate the next trial. Training continued until the seal had reached the
195 criterion of seven consecutive sessions with overall accuracy above 80%, at which point
196 testing began. During testing, stimuli were presented in blocks of 50 trials. Depending on
197 how quickly trials were completed (due to breaks for environmental noise), each test session
198 would consist of either 50 or 100 trials (mean session duration = 47 minutes, $SD \pm 7$). If the

A SEAL GENERALIZES CALL CLASSES

199 seal's response overlapped part of the played stimulus, the seal was not reinforced, regardless
200 of whether she responded correctly or not (this occurred in 136 of the 1,182 training and 72
201 of the 700 test trials). As call type could still be determined, the seal's calls were still used
202 when measuring its accuracy in the response.

203 **Sound Stimuli and Acoustic Analysis**

204 Training stimuli were composed of Zola's own previously recorded vocalizations in
205 sets of 10 to 20 novel exemplars per session. The sounds were played randomly with one
206 exception; if Zola responded incorrectly, the sound was replayed up to three times until she
207 produced the correct response. If after three replays Zola had not responded correctly we
208 moved on to the next call. This 'self-correction' procedure was only used during training
209 trials. In total, 212 different calls (106 'growls' and 106 'moans') were used throughout
210 training.

211 Test stimuli consisted of 140 novel calls, 70 produced by Zola and 70 from the two
212 juvenile seals studied by Shapiro et al. (2004), Kylie and Oscar, who were unknown to Zola.
213 All calls were between 0.7 and 1.2 seconds in duration. During each trial, one of the 140
214 novel call stimuli was played. Stimuli were presented in blocks of 50 trials (five 'growls' and
215 five 'moans'), presented five times per stimulus in random order assigned by a Gellerman
216 series (Gellerman, 1933). Thus, Zola was tested with 14 test blocks (700 trials). Frequency
217 and time parameters were measured using Avisoft-Saslab Pro 5.02.04 sonogram software.
218 Parameters were chosen based on previous studies examining grey seal vocal repertoires
219 (Asselin, Hammill, & Barrette, 1993; McCulloch, Pomeroy, & Slater, 1999). See Table 1 for
220 a list of measured parameters and their definitions.

221 **Results**

222 **Call Type Classification**

A SEAL GENERALIZES CALL CLASSES

223 All statistical analyses were conducted using R 3.0.1. Two independent human
224 observers classified Zola's responses as 'growls' or 'moans' post hoc. At this time, the
225 observers classified only the seal's response; they were unaware of which stimulus was
226 played during each trial. To confirm that these calls were classified appropriately, Cohen's
227 Kappa was used to assess inter-observer reliability, with high agreement (99.7% agreement,
228 Cohen's Kappa = 0.9547, SE \pm 0.0109). K-means cluster analysis (k = 2) was conducted
229 using the frequency and time parameters measured to evaluate human classification. All calls
230 grouped by K-means were in agreement (100%) with one of the two human's categorization,
231 which was also in complete agreement (100%) with the experimenter who reinforced the seal
232 during testing. Thus, this grouping was deemed objective and used to score the seal's
233 accuracy for the remaining statistical analysis.

234 Training

235 The seal's accuracy during training sessions (all sessions until criteria of seven
236 consecutive sessions >80% correct was reached) were plotted as percentage correct per
237 session by call type (Figure 2). Zola reached criterion after 1,182 trials across 18 sessions.
238 This was comparable to Kylie (~1,650 trials, 47 sessions) and Oscar's (~800 trials, 23
239 sessions) performance (Shapiro et al., 2004). A nonlinear least squares (NLS) model was
240 used to examine how the subject's overall accuracy changed over the learning period by date
241 and call type. As session length throughout training was dependent upon the seal's
242 performance, the number of trials per session varied and the model was weighted by number
243 of trials per day. The seal's overall accuracy was plotted with a polynomial trendline fitted
244 for the best fit model, determined using the lowest Akaike Information Criterion (AICc) score
245 tested for all orders up to seven (Figure 2).

246 Testing

A SEAL GENERALIZES CALL CLASSES

247 For test sessions, the seal's accuracy was plotted as percentage correct by blocks of
248 50 trials and compared between her own and the calls of the two unknown seals (Figure 3).
249 Overall Zola responded with 93.7% accuracy to her own calls and with 87.2% accuracy to
250 Kylie and Oscar's calls. This performance was significantly above chance ($\chi^2(1, N=700) =$
251 426.05, $p < 0.001$).

252 To determine if alternative parameters might have affected the seal's test
253 performance, we used a generalized linear model (GLM). Accuracy was predicted using a
254 binomial distribution and logit link function with the following covariates: date, trial number,
255 signal exposure (the number of times the seal had been exposed to stimuli; each test signal
256 was played five times), call source (Zola, Kylie and Oscar's calls), and call type ('growl'
257 versus 'moan'). All possible combinations were tested including interactions between date,
258 call source, and call type. The best models were selected using the AICc score; all models
259 within 10 AICc points of the best model were used for model averaging (the model's selected
260 for averaging are shown in supplementary material 1). The best models were averaged using
261 modelavg (package MuMIn, version 1.9.13: Barton, 2013) to determine the final model. The
262 model indicated that Zola's accuracy throughout testing significantly improved across
263 sessions (Table 2). However, her accuracy decreased with signal exposure across sessions.
264 Each test signal was played five times, and Zola had significantly higher accuracy during the
265 first exposures, and lower accuracy for the last exposure. This significant decrease in
266 accuracy across exposures occurred when hearing her own calls, but not for Kylie and
267 Oscar's calls (Table 2 and supplementary material 2). However, the degree to which these
268 variables affect the model is very small. The covariates only explain a small proportion
269 (7.5%) of the variance (Nagelkirke's pseudo $r^2 = 0.075$).

270 **FAMD**

A SEAL GENERALIZES CALL CLASSES

271 Descriptive statistics (mean and standard deviation) for measured acoustic parameters
272 by call type are shown in Table 3. Factor analysis for mixed data (FAMD) was used to
273 further examine the variability between calls and identify possible cues the seal could have
274 used to distinguish between call types (package FactoMineR, version 1.26: Husson, Josse,
275 Le, & Mazet, 2014). FAMD was chosen because it allowed both continuous and categorical
276 parameters to be used to group calls similarly to K-means cluster analysis (Ding & He, 2004).
277 Calls were analysed both by call type ('growl' and 'moan') and by individual (Zola, Kylie,
278 and Oscar). These methods allowed us to determine what parameters explained the most
279 variation between call types across individuals, and those that explained differences between
280 individuals.

281 The Kaiser-Meyer-Olkin (KMO) measure verified that sample size was adequate with
282 a KMO = 0.78. Bartlett's test of sphericity ($\chi^2(903) = 89533.15, p < 0.001$) indicated that
283 correlations between variables were sufficiently large to conduct FAMD. Dimensions were
284 plotted against each other to examine clustering by call type and individual. For call type,
285 'growls' and 'moans' were clearly clustered for the first three dimensions, which explained
286 64.16% of the data's variance (Figure 4a). Calls were clustered by individual for the first two
287 dimensions, which explained 53.29% ('growls') and 40.9% ('moans') of the data's variance
288 (Figure 4b and 4c).

289 To identify what parameters explained the variance between call types, factor
290 loadings for the first 3 dimensions were examined (supplementary material 3). Based on the
291 loading weights these related to noise content, frequency and duration parameters. Call
292 clustering suggested that 'growls' scored high on noise content and frequency but low on
293 duration dimensions. 'Moans' were calls that scored low on noise content and frequency but
294 high on duration dimensions. The first dimension was also heavily weighted by categorical
295 parameters; 'growls' had no fundamental frequency, and scored low on signal to noise

A SEAL GENERALIZES CALL CLASSES

296 categories, and ‘moans’ had a fundamental frequency and scored high on signal to noise
297 categories. When examining parameters that explained variance between individuals, factor
298 loadings for the first 2 dimensions were examined (supplementary material 3). Based on the
299 loading weights, these components related to frequency parameters. Call clustering by
300 individual suggested that for ‘growls’, individuals differed in peak frequency parameters. For
301 ‘moans’, both fundamental and peak frequency differed between individuals, however
302 fundamental frequency explained more of the individual variation (supplementary material
303 3).

Discussion

305 This study demonstrated that a grey seal was clearly capable of categorizing and
306 generalizing between call classes, contrasting with Shapiro et al.’s (2004) findings that grey
307 seals were unable to generalize novel stimuli. This disparity may have been due to
308 methodological differences. Shapiro et al.’s (2004) seals were exposed to a small set of
309 training stimuli and then tested with novel stimuli. As the seals were trained with such a
310 small set they may have learned to respond to each sound as an individual cue, rather than to
311 generalize to new stimuli. Other studies utilizing small training sets have been similarly
312 unsuccessful at training call type categorization of novel stimuli in other species (Le Prell et
313 al., 2002; Vergara, 2011). In generalization tasks using other modalities, training set size has
314 been found to impact learning speed and accuracy (e.g., Katz & Wright, 2006; Schrier &
315 Brady 1987; Wasserman & Bhatt 1992; Wright et al., 1988; Wright et al., 2003). Thus, Kylie
316 and Oscar’s inability to generalize may be attributed to the training with small number of
317 exemplars; through our use of a variety of stimuli Zola was required to generalize from the
318 start of training, and was thereby successful when tested with novel stimuli.

319 The duration of stimuli used by Shapiro et al. (2004) also varied between training and
320 testing sets, which may have prevented the seal’s successful classification of test stimuli if

A SEAL GENERALIZES CALL CLASSES

321 absolute duration was used as a cue. Zola was tested with calls of standardized duration,
322 which may have facilitated her success. Interestingly, although the duration of all calls fell
323 within a small range (0.7 and 1.2 seconds), call types could still be separated by duration.
324 Additionally, Zola was two months older than the animals studied by Shapiro et al. (2004),
325 but it seems unlikely that this would have made a difference as she appeared to learn the task
326 at comparable rates to Shapiro et al.'s (2004) animals.

327 Zola performed significantly better during the first exposures to individual signals in
328 the tests than later on, supporting further that the seal generalized signal type rather than
329 learned items individually. The decrease in accuracy for later signal exposures may be
330 attributed to fatigue; the later presentations of each signal would occur near the end of each
331 block and long sessions may have impacted accuracy. Interestingly, the average duration of
332 sessions consisting of both 50 and 100 trials were comparable (overall mean duration = 47
333 minutes, ± 7). The similar overall session length for sessions consisting of either one or two
334 blocks may be attributed to variable inter-trial intervals; during some sessions, background
335 noise was low and several trials were quickly completed, allowing for two blocks to be
336 finished within one session. However, on days with varying background noise the inter-trial
337 interval was sometimes longer as trials did not resume until background noise was reduced,
338 only allowing for one block to be completed. Future studies could address effects of fatigue
339 and trial variability by using shorter test blocks with fixed inter-trial intervals.

340 The factor analysis indicated that call types could be classified using noise content,
341 frequency, and duration parameters whereas individuals could be separated using frequency
342 parameters. 'Growls' had a higher maximum call frequency than 'moans', but lower peak
343 frequency (Table 3). Within 'growls', individuals were separated by peak frequency,
344 whereas for 'moans' although peak frequency contributed, fundamental frequency explained
345 more of the individual variation (Figure 5 and supplementary material 3). Thus, the seal

A SEAL GENERALIZES CALL CLASSES

346 could have used different frequency parameters to provide information about both call type
347 and caller identity. These results are also comparable to McCulloch et al. (1999) findings
348 that wild grey seal pup calls could be individually identified using duration and frequency
349 parameters. Interestingly, although the calls were found to be individually distinctive,
350 mothers did not differentially respond to playbacks of their own versus unknown pups in
351 Scotland (McCulloch et al., 1999). This was not the case at a different seal colony, Sable
352 Island, where grey seal mothers were found to respond to playback of their own pups
353 (McCulloch & Boness, 2000). Although grey seal vocalizations appear to provide caller
354 information, the biological significance is unclear.

355 Interestingly, one of Shapiro et al.'s (2004) seals tended to reply to any novel stimulus
356 with a 'growl'. It is possible that the seal responded to the novel stimuli as if presented with
357 an intruding conspecific. Zola similarly responded to 'growls' by 'growling' even in the first
358 session of training, suggesting that initially her response may also be a natural reply to the
359 'growl' of a conspecific rather than a learned response. Later in the sessions, her 'growling'
360 response deteriorated, most likely due to a habituation effect in the playback context, before
361 she successfully replied to 'growls' by 'growling' again after a learning period (Fig 2). The
362 initial 'growling' in our and Shapiro et al.'s (2004) study suggests that 'growls' may be
363 aggressive or defensive signals when challenged by a conspecific. While our study only
364 examined calls produced in an artificial training context and can thus only hypothesize as to
365 the function of these call types, future studies could examine the biological significance of
366 these calls in wild animals using playback studies.

367 It is unknown what information Zola used for discriminating between call types. She
368 may have used a single parameter to classify the calls, or a combination of multiple
369 parameters. Frequency parameters between call types and individual may be more reliable in
370 combination with other call features. Although parameters salient to human researchers

A SEAL GENERALIZES CALL CLASSES

371 were measured, it is also possible other unconsidered cues were used for discrimination by
372 the seal. Features relevant to humans may or may not also be relevant to nonhuman species,
373 which is a commonly acknowledged problem when examining animal communication (e.g.,
374 Harley, 2008). However, the fact that Zola performed successfully in these trials
375 demonstrated that generalization across different versions of the same call type is possible in
376 this species. At this time, it is unclear whether Zola used an exemplar based strategy for this,
377 where features that best predicted group membership across multiple learned examples were
378 used to classify novel calls, or prototype based discrimination, which is based on similarity to
379 a ‘best’ example of each call type. Future research could investigate these issues by altering
380 training stimuli systematically.

381 Further studies on learning, production and perception capabilities and limitations of
382 vocal behaviour would be valuable for pinnipeds given their interesting vocal learning skills
383 (Janik, 2006) and their dynamic social structures (Ruddell, Twiss, & Pomeroy, 2006).

384

385 **Acknowledgements**

386 We would like to thank Simon Moss, Matt Bivins, Ryan Milne, Alicia Widmer, and all
387 of the student volunteers (particularly Sian Tarrant and Helen Hiley) at our facility for their
388 help during this study. We also thank the anonymous reviewers for their insight. This study
389 was funded by the Natural Environment Research Council of the UK, grant no.
390 NE/I024682/1. It was conducted under Home Office licence number 60/3303.

391

392

393

394

395

396

References

- 397 Asselin, S., Hammill, M.O., & Barrette, C. (1993). Underwater vocalizations of ice breeding
 398 grey seals. *Canadian Journal of Zoology*, *71*, 2211-2219. doi: 10.1139/z93-310
- 399 Audacity Team (2012). Audacity, Version 1.3. <http://audacity.sourceforge.net/>
- 400 Barton, K. (2013). MuMIn: Multi-model inference. R package version 1.9.13.
 401 <http://CRAN.R-project.org/package=MuMIn>
- 402 Beecher, M.D., Campbell, S.E., & Burt, J.M. (1994). Song perception in the song sparrow:
 403 Birds classify by song type but not by singer. *Animal Behaviour*, *47*, 1343-1351. doi:
 404 10.1006/anbe.1994.1182
- 405 Bloomfield, L. L., Sturdy, C. B., Phillmore, L. S., & Weisman, R. G. (2003). Open-ended
 406 categorization of chick-a-dee calls by black-capped chickadees (*Poecile atricapilla*).
 407 *Journal of Comparative Psychology*, *117*, 290 - 301. doi: 10.1037/0735-7036.117.3.290
- 408 Brooks, D.I., & Cook, R.G. (2010). Chord discrimination by pigeons. *Music Perception: An*
 409 *Interdisciplinary Journal*, *27*, 183-196. doi: 10.1525/mp.2010.27.3.183
- 410 Candiotti, A., Zuberbühler, K., & Lemasson, A. (2013). Voice discrimination in four
 411 primates. *Behavioural Processes*, *99*, 67-72. doi: 10.1016/j.beproc.2013.06.010
- 412 Cäsar, C., Zuberbühler, K., Young, R. J., & Byrne, R. W. (2013). Titi monkey call sequences
 413 vary with predator location and type. *Biology Letters*, *9*, 20130535. doi:
 414 10.1098/rsbl.2013.0535
- 415 Ding, C., & He, X. (2004). K-means clustering via principle components analysis.
 416 *Proceedings of the Twenty-First International Conference on Machine Learning*. ACM,
 417 pg. 29.
- 418 Gellerman, L.W. (1933). Chance orders of alternating stimuli in visual discrimination
 419 experiments. *The Pedagogical Seminary and Journal of Genetic Psychology*, *42*, 206-
 420 208. doi:10.1080/08856559.1933.10534237

A SEAL GENERALIZES CALL CLASSES

- 421 Ghazanfar, A. A., Flombaum, J. I., Miller, C. T., & Hauser, M. D. (2001). The units of
422 perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus*
423 *oedipus*): Playback experiments with long calls. *Journal of Comparative Physiology A*,
424 *187*, 27-35. doi: 10.1007/s003590000173
- 425 Greene, E., & Meagher, T. (1998). Red squirrels, *Tamiascirus hudsonicus*, produce
426 predator-class alarm calls. *Animal Behaviour*, *55*, 511-518. doi:
427 10.1006/anbe.1997.0620
- 428 Harley, H. E. (2008). Whistle discrimination and categorization by the Atlantic bottlenose
429 dolphin (*Tursiops truncatus*): A review of the signature whistle framework and a
430 perceptual test. *Behavioural Processes*, *77*, 243-268. doi: 10.1016/j.beproc.2007.11.002
- 431 Helverson, D.V. (2004). Object classification by echolocation in nectar feeding bats: Size-
432 independent generalization of shape. *Journal of Comparative Physiology A*, *190*, 515-
433 521. doi: 10.1007/s00359-004-0492-9
- 434 Hoeschele, M., Cook, R. G., Guillette, L. M., Brooks, D. I., & Sturdy, C. B. (2012). Black-
435 capped chickadee (*Poecile atricapillus*) and human (*Homo sapiens*) chord
436 discrimination. *Journal of Comparative Psychology*, *126*, 57-67. doi:
437 10.1037/a0024627
- 438 Husson, F., Josse, J., Le, S., & Mazet, J. (2014). FactoMineR: Multivariate exploratory data
439 analysis and data mining with R. R package version 1.26. [http://CRAN.R-](http://CRAN.R-project.org/package=FactoMineR)
440 [project.org/package=FactoMineR](http://CRAN.R-project.org/package=FactoMineR)
- 441 Janik, V. M., & Slater, P. J. (2000). The different roles of social learning in vocal
442 communication. *Animal Behaviour*, *60*, 1-11. doi: 10.1006/anbe.2000.1410
- 443 Janik, V. M. (2006). Communication in marine mammals. In K. Brown (Ed.), *Encyclopedia*
444 *of language and linguistics (2nd edition)*. Vol. 2, (pp. 646-654). Oxford: Elsevier.

A SEAL GENERALIZES CALL CLASSES

- 445 Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: evidence for
446 "fast mapping". *Science*, *304*, 1682-1683. doi: 10.1126/science.1097859
- 447 Katz, J.S., & Wright, A.A. (2006). Same-different abstract concept learning by pigeons.
448 *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 80-86. doi:
449 10.1037/0097-7403.32.1.80
- 450 Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory
451 sensitivities conducive to the perception of speech-sound categories. *The Journal of the*
452 *Acoustical Society of America*, *70*, 340-349. doi: 10.1121/1.386782
- 453 Le Prell, C.G., Hauser, M.D., & Moody, D.B. (2002). Discrete or graded variation within
454 rhesus monkey screams? Psychophysical experiments on classification. *Animal*
455 *Behaviour*, *63*, 47-62. doi: 10.1006/anbe.2001.1888
- 456 Lind, H., Dabelsteen, T., & McGregor, P. K. (1996). Female great tits can identify mates by
457 song. *Animal Behaviour*, *52*, 667-671. doi: 10.1006/anbe.1996.0211
- 458 Maros, K., Pongrácz, P., Bárdos, G., Molnár, C., Faragó, T., & Miklósi, A. (2008). Dogs can
459 discriminate between barks from different situations. *Applied Animal Behaviour*
460 *Science*, *114*, 159-167. doi: 10.1016/j.applanim.2008.01.022
- 461 McCulloch, S., & Boness, D J. (2000). Mother-pup vocal recognition in grey seals
462 (*Halichoerus grypus*) of Sable Island, Nova Scotia, Canada. *Journal of Zoology*, *251*,
463 449-455. doi: 10.1111/j.1469-7998.2000.tb00800.x
- 464 McCulloch, S., Pomeroy, P.P., & Slater, P.J.B. (1999). Individually distinctive pup
465 vocalizations fail to prevent allo-suckling in grey seals. *Canadian Journal of Zoology*,
466 *77*, 616-723. doi: 10.1139/z99-023
- 467 Mercado, E., Killebrew, D.A., Pack, A.A., Mácha, I.V.B., & Herman, L.M. (2000).
468 Generalization of 'same-different' classification abilities in bottlenosed dolphins.
469 *Behavioural Processes*, *50*, 79-94. doi: 10.1016/S0376-6357(00)00100-5

A SEAL GENERALIZES CALL CLASSES

- 470 Miller, E. H., & Murray, A. V. (1995). Structure, complexity, and organization of
471 vocalizations in harp seal (*Phoca groenlandica*) pups. In P.E. Nachtigall, J.A. Thomas,
472 & R.A. Kastelein (Eds.), *Sensory systems of aquatic mammals* (pp. 237-264).
473 Netherlands: De Spil Publishers.
- 474 Neiworth, J.J., & Wright, A.A. (1994). Monkeys (*Macaca mulatta*) learn category matching
475 in a nonidentical same-different task. *Journal of Experimental Psychology: Animal*
476 *Behaviour Processes*, *20*, 429-435. doi: 10.1037/0097-7403.20.4.429
- 477 Pepperberg, I.M., & Brezinsky, M.V. (1991). Acquisition of a relative class concept by an
478 African gray parrot (*Psittacus erithacus*): Discrimination based on relative size. *Journal*
479 *of Comparative Psychology*, *105*, 286-294. doi: 10.1037/0735-7036.105.3.286
- 480 Pepperberg, I.M. (1996). Categorical class formation by an African grey parrot (*Psittacus*
481 *erithacus*). In T.R. Zentall, & P.M. Smeets (Eds.), *Advances in psychology: Stimulus*
482 *class formation in humans and animals* (pp. 71-90). Amsterdam: Elsevier. doi:
483 10.1016/S0166-4115(06)80104-4
- 484 Pepperberg, I.M. (1999). *The Alex studies: Cognitive and communicative abilities of grey*
485 *parrots* (pp. 434). Cambridge, MA: Harvard University Press.
- 486 Porter, D., & Neuringer, A. (1984). Music discriminations by pigeons. *Journal of*
487 *Experimental Psychology: Animal Behavior Processes*, *10*, 138. doi: 10.1037/0097-
488 7403.10.2.138
- 489 Range, F., Aust, U., Steurer, M., & Huber, L. (2008). Visual categorization of natural stimuli
490 by domestic dogs. *Animal Cognition*, *11*, 339-347. doi: 10.1007/s10071-007-0123-2
- 491 Roitblat, H.L., Penner, R.H., & Nachtigall, P.E. (1990). Matching-to-sample by an
492 echolocating dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology:*
493 *Animal Behavior Processes*, *16*, 85-95. doi: 10.1037/00977403.16.1.85

A SEAL GENERALIZES CALL CLASSES

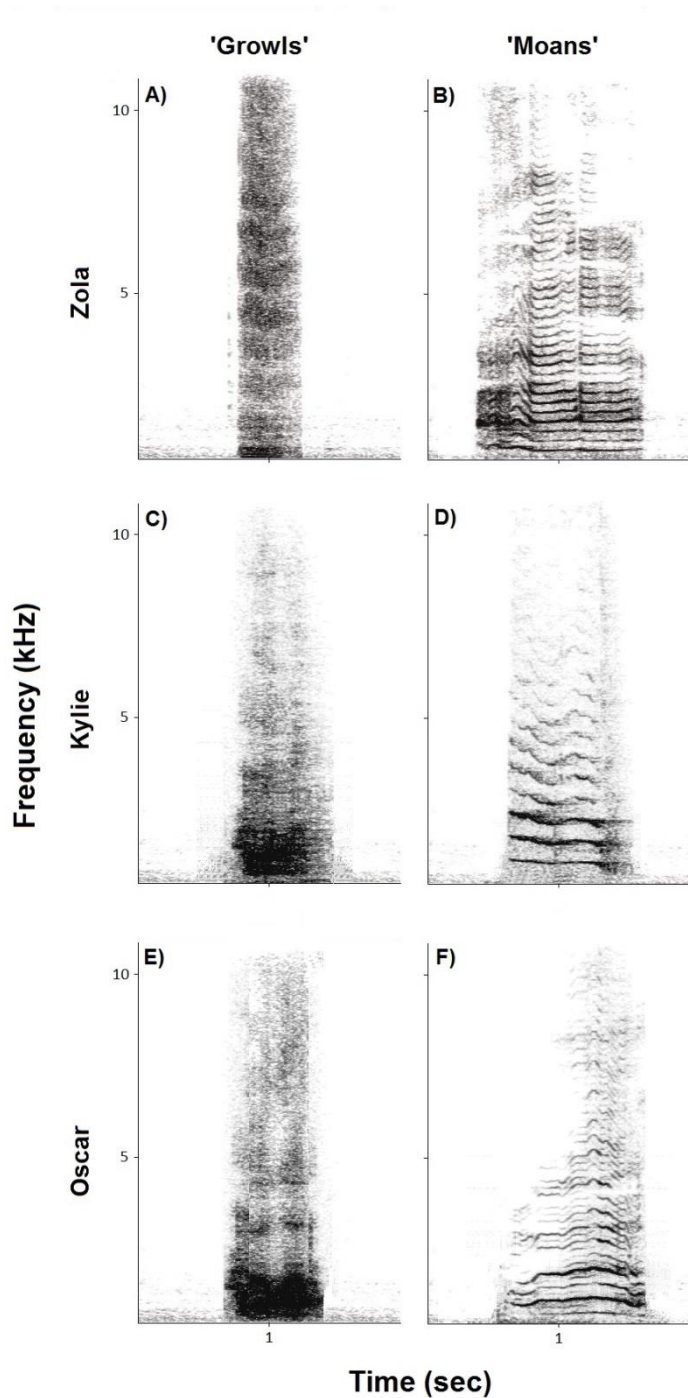
- 494 Ruddell, S.J.S., Twiss, S.D., & Pomeroy, P.P. (2007). Measuring opportunity for sociality:
495 quantifying social stability in a colonially breeding phocid. *Animal Behaviour*, *74*,
496 1357-1368. doi: 10.1016/j.anbehav.2007.01.024
- 497 Savage-Rumbaugh, E.S., Murphy, J., Sevcik, R.A., Brakke, K.E., Williams, S.L., Rumbaugh,
498 D.M., & Bates, E. (1993). Language comprehension in ape and child. *Monographs of*
499 *the Society for Research in Child Development*, *58*, 1-221.
- 500 Schrier, A.M., & Brady, P.M. (1987). Categorization of natural stimuli by monkeys (*Macaca*
501 *mulatta*): Effects of stimulus set size and modification of exemplars. *Journal of*
502 *Experimental Psychology: Animal Behavior Processes*, *13*, 136-143. doi:
503 10.1037/0097-7403.13.2.136
- 504 Seyfarth, R.M., Cheney, D.L., & Marler, P. (1980). Monkey responses to three different
505 alarm calls: Evidence of predator classification and semantic communication. *Science*,
506 *210*, 801-803. doi: 10.1126/science.7433999
- 507 Shapiro, A.D., Slater, P.J.B., & Janik, V.M. (2004). Call usage learning in gray seals
508 (*Halichoerus grypus*). *Journal of Comparative Psychology*, *118*, 447-454. doi:
509 10.1037/0735-7036.118.4.447
- 510 Soltis, J., Leong, K., & Savage, A. (2005). African elephant vocal communication I:
511 Antiphonal calling behaviour among affiliated females. *Animal Behaviour*, *70*, 579-
512 587. doi: 10.1016/j.anbehav.2004.11.015
- 513 Vergara, V. (2011). Acoustic communication and vocal learning in belugas (*Delphinapterus*
514 *leucas*). pp. 153, PhD thesis: The University of British Columbia.
- 515 Wasserman, E.A., & Bhatt, R.S. (1992). Conceptualization of natural and artificial stimuli by
516 pigeons. In W.K. Honig, & J.G. Fetterman (Eds.), *Cognitive aspects of stimulus control*
517 (pp. 203-223). Hillsdale, NJ: Erlbaum.

A SEAL GENERALIZES CALL CLASSES

- 518 Wasserman, E.A., Zentall, T.R., & Cook, R.G. (2006). Relational discrimination learning in
519 pigeons. In E.A. Wasserman & T.R. Zentall (Eds.), *Comparative cognition* (pp. 307-
520 324). New York, NY: Oxford University Press.
- 521 Watanabe, S. (2001). Van Gogh, Chagall and pigeons: Picture discrimination in pigeons and
522 humans. *Animal Cognition*, 4, 147-151. doi: 10.1007/s100710100112
- 523 Weary, D. M., & Krebs, J. R. (1992). Great tits classify songs by individual voice
524 characteristics. *Animal Behaviour*, 43, 283-287. doi: 10.1016/S0003-3472(05)80223-4
- 525 Wright, A.A., Cook, R.G., Rivera, J.J., Sands, S.F., & Delius, J.D. (1988). Concept learning
526 by pigeons: Matching to sample with trial-unique video picture stimuli. *Animal*
527 *Learning & Behavior*, 16, 436-444. doi: 10.3758/BF03209384
- 528 Wright, A.A., Rivera, J.J., Katz, J.S., & Bachevalier, J. (2003). Abstract-concept learning and
529 list-memory processing by capuchin and rhesus monkeys. *Journal of Experimental*
530 *Psychology: Animal Behavior Processes*, 29, 184-198. doi: 10.1037/0097-
531 7403.29.3.184
- 532
- 533
- 534
- 535
- 536
- 537
- 538
- 539
- 540
- 541

A SEAL GENERALIZES CALL CLASSES

542 Figure 1. Spectrographic examples of call types from the three seals. ‘Growls’ are broadband,
543 noisy calls, and ‘moans’ are tonal, periodic calls with a harmonic structure. Example ‘growls’
544 are shown from A) Zola, C) Kylie and E) Oscar and ‘moans’ from B) Zola, D) Kylie and F)
545 Oscar. Spectrograms created in Avisoft-SASlab Pro (FFT size: 2048, frequency resolution:
546 46.87 Hz, time resolution: 10.66 ms, weighting function: hamming, window width: 100%).

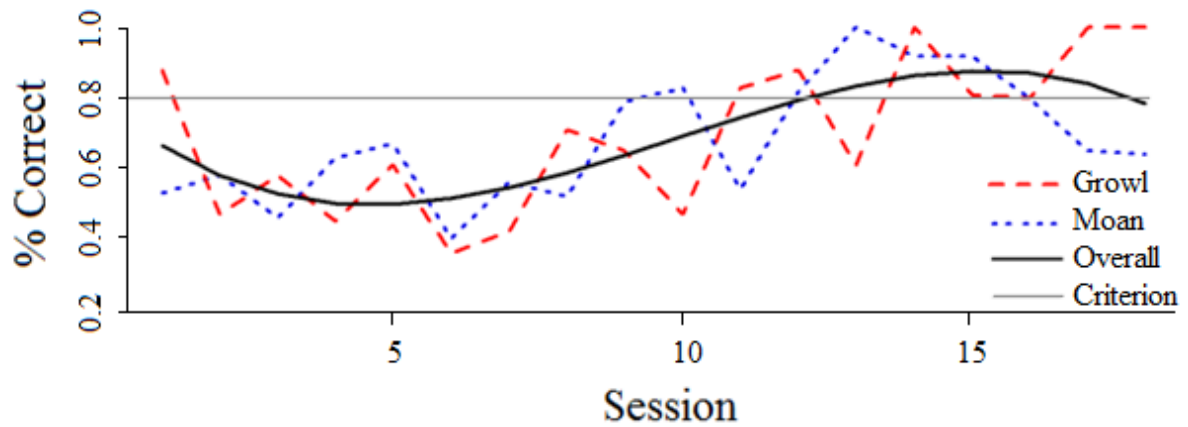


547

548

A SEAL GENERALIZES CALL CLASSES

549 Figure 2. Performance accuracy by session for Zola's training period. Accuracy (percentage
550 correct) is shown by call type ('growls' and 'moans'). A nonlinear least squares model was
551 used to determine the best fit polynomial trendline (order 3) for overall performance,
552 weighted by the number of trials per session. Training continued until a criterion of 7
553 consecutive sessions with >80% accuracy for 'growls' and 'moans' combined was met.



554

555

556

557

558

559

560

561

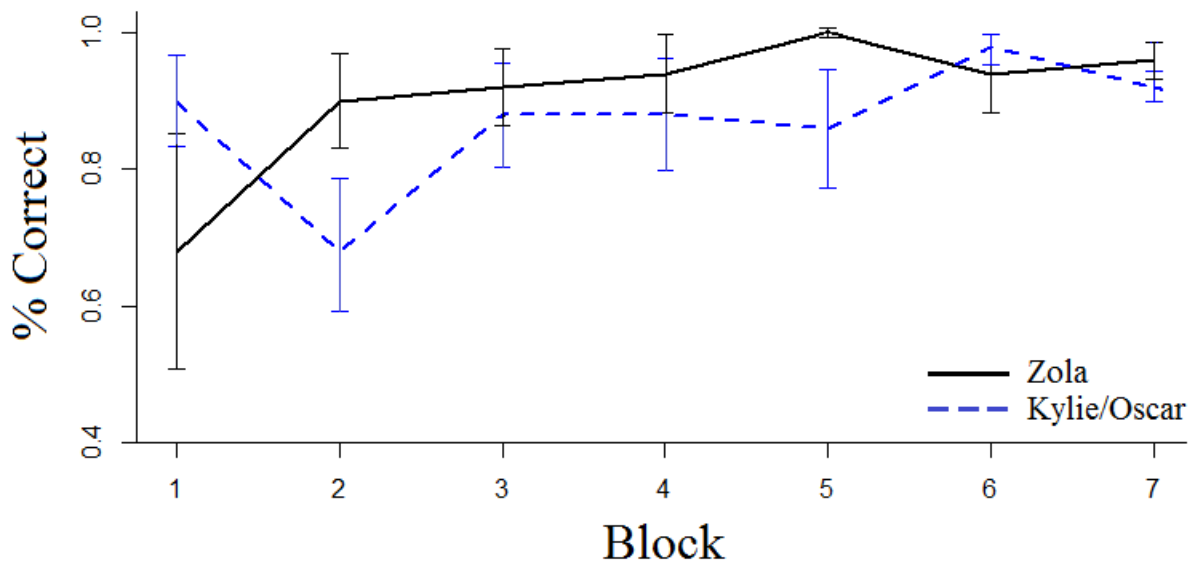
562

563

564

A SEAL GENERALIZES CALL CLASSES

565 Figure 3. Test session accuracy by blocks of 50 trials. Accuracy (percentage correct) and
566 standard deviation is shown, separated by test stimuli source (Zola: solid line, Kylie and
567 Oscar, whose calls were taken from Shapiro et al., 2004: dashed line).



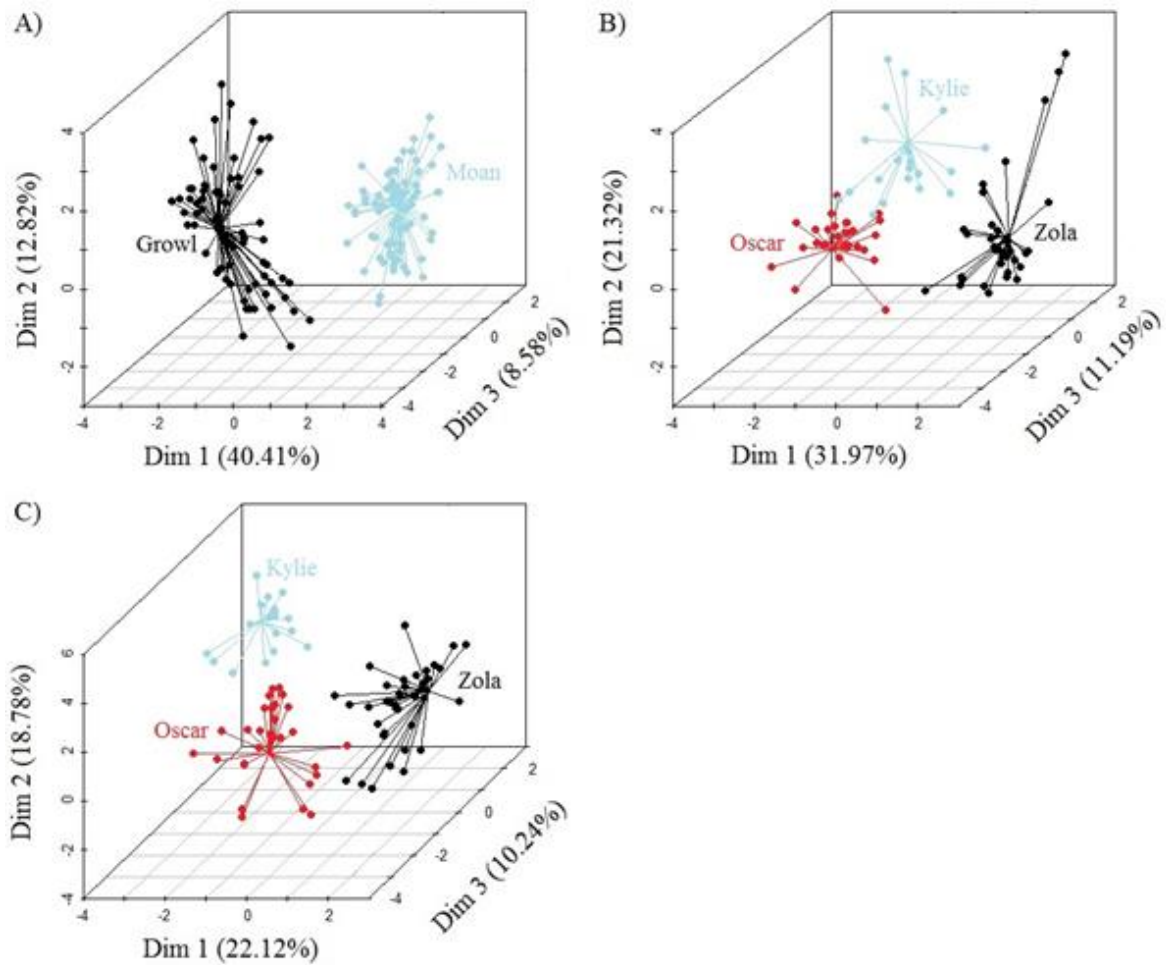
568

569

570

A SEAL GENERALIZES CALL CLASSES

571 Figure 4. 3D scatterplots of components 1 through 3 from factor analysis for mixed data
 572 (FAMD). The proportions of the data's variance explained by each component are shown in
 573 parentheses. 4a) Calls from all three seals clustered by call type. 4b) 'Growls' clustered by
 574 individual, 4c) 'Moans' clustered by individual. See supplementary material 3 for table of
 575 factor loadings showing how variables loaded onto the dimensions.



576

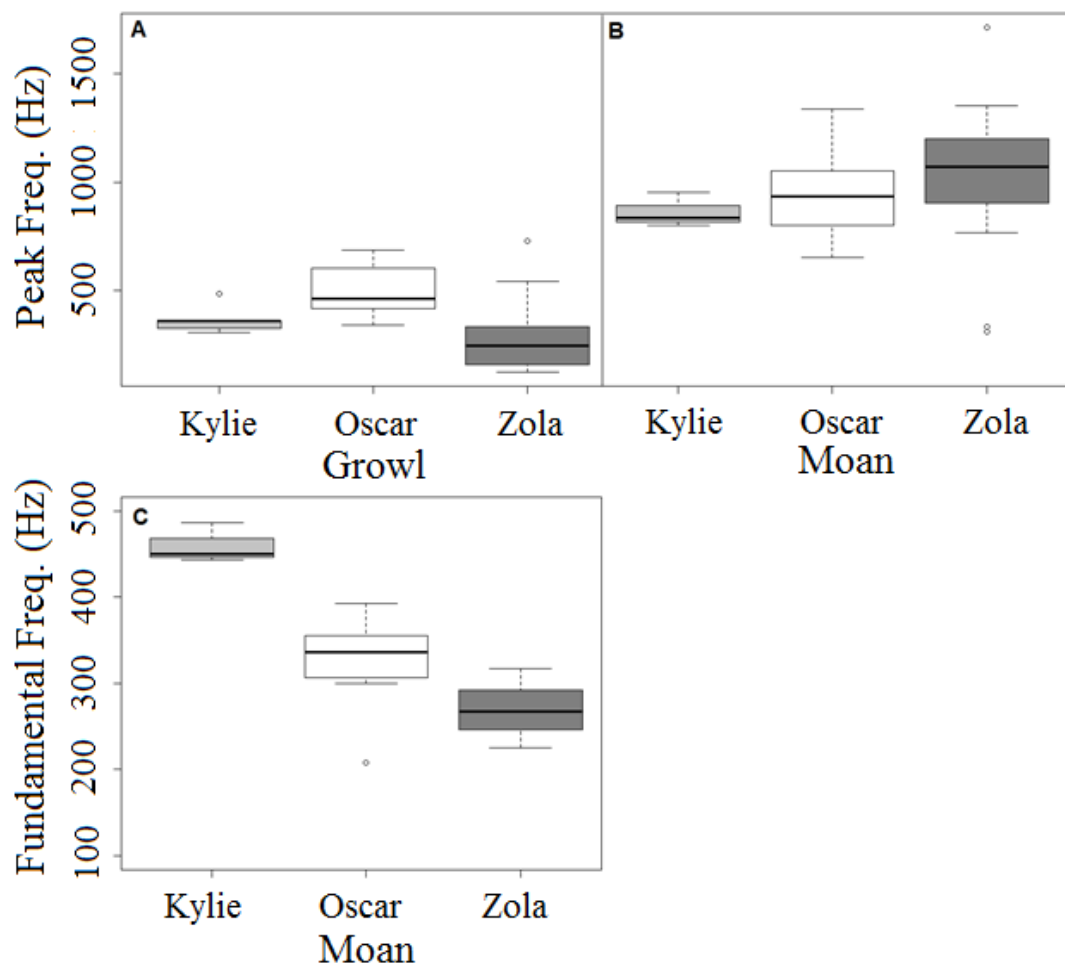
577

578

579

A SEAL GENERALIZES CALL CLASSES

580 Figure 5. Tukey's boxplots for test stimuli by caller showing the peak frequency for A)
 581 'growls' and B) 'moans', and fundamental frequency for C) 'moans'. Fundamental frequency
 582 for 'growls' is not shown as 'growls' are broadband calls without a periodic structure.



583

584

A SEAL GENERALIZES CALL CLASSES

585 Table 1. Definitions for measured parameters. All parameters were measured from
 586 spectrograms (FFT size: 2048 frequency resolution: 46.87 Hz, time resolution: 10.66 ms.).

Parameter	Definition
Duration	Time from the beginning to end of the call, within 35 dB of the spectrums maximum peak.
Peak Frequency	The frequency with the highest amplitude measured at the beginning, middle, end, and maximum (i.e., the frequency with the highest amplitude) of the whole call.
Fundamental Frequency	Measured only for tonal, harmonic calls. Measured at the centre of the call as the lowest integer multiple of corresponding amplitude peaks.
Maximum Frequency	The highest frequency reached at any time point within 20 dB of the spectrums maximum peak.
Wiener Entropy	The ratio of the geometric mean to the arithmetic mean of the spectrum. This parameter is closer to zero for pure-tone signals and closer to one for random noise. The signal was measured at the centre of the call, and compared to background noise around the call.
Harmonic to Noise ratio (HNR)	The dB ratio between the harmonic and non-harmonic energy. The ratio is averaged for each frequency bin (46.87 Hz), for a 10.66 ms section measured at the centre of the call.
Spectral Richness (SR)	A measure of broadband noise and harmonic spacing as defined by Miller & Murray (1995). Calls are rated on a scale of 0 to 3 based on separation between harmonic bands (H) and the width of the fundamental frequency band (F). F is measured by subtracting the lowest from the highest frequency of the fundamental at the band's widest point. H is measured by subtracting the highest frequency of the fundamental from the lowest frequency of the second harmonic at the centre of the call. A score of 0 is applied to broadband signals, 1 if $H > F$, 2 if $H = F$, and 3 $H < F$.

587

588

589

590

591

592

A SEAL GENERALIZES CALL CLASSES

593 Table 2. Final generalized linear model (GLM) selected using model averaging. The models
 594 used for averaging are shown in supplementary material 1. Log adjusted model weights, 95%
 595 confidence interval, and P-value are shown.

	Model Estimate e^{β}	95% Confidence Interval		P-Value
		Lower	Higher	
Intercept	0.421	0.028	0.950	0.848
Call Source	0.736	0.220	0.965	0.381
Date	0.643	0.517	0.751	0.026
Signal Exposure	0.500	0.299	0.701	0.999
Date*Signal Exposure	0.470	0.447	0.493	0.009
Call Source*Signal Exposure	0.630	0.502	0.741	0.047

596

597

598

599

600

601

602

603

604

605

A SEAL GENERALIZES CALL CLASSES

606 Table 3. Mean and standard deviation for measured acoustic parameters by call type. Values
607 averaged for 'growls' and 'moans' across individuals.

	'Growls'	'Moans'
Duration (secs)	0.85 ± 0.12	0.92 ± 0.12
Fundamental Frequency (Hz)	N/A	357 ± 141
Peak Frequency (Hz)	522 ± 358	874 ± 300
Maximum Frequency (Hz)	16512 ± 3221	12989 ± 2431
Weiner Entropy	0.45 ± 0.18	0.28 ± 0.09
Harmonic to Noise Ratio	16 ± 5.7	23.5 ± 5.5

608

609

610

611

612

613

614

615

616

617

618

619