



# Memory for own behaviour in pinnipeds

Simeon Quirinus Smeele<sup>1</sup> · Kirstin Anderson Hansen<sup>1,2</sup> · Sara Torres Ortiz<sup>1,3</sup> · Fredrik Johansson<sup>4</sup> · Jakob Højer Kristensen<sup>4</sup> · Josefin Larsson<sup>4</sup> · Ursula Siebert<sup>2</sup> · Magnus Wahlberg<sup>1</sup>

Received: 29 October 2018 / Revised: 9 June 2019 / Accepted: 18 June 2019 / Published online: 25 June 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

Pinnipeds are aquatic predators feeding on a vast range of prey, and their social behaviour differs greatly between species (from extreme polygyny in some sea lions to monogamy in some true seals). It has been hypothesised that the foraging and social complexity of their lifestyle should drive the evolution of their cognitive abilities. To investigate how aware pinnipeds are of their own behaviour, a grey seal (*Halichoerus grypus*), two harbour seals (*Phoca vitulina*) and four South American sea lions (*Otaria flavescens*) were trained to repeat their own behaviour on command. Three already trained behaviours were used, and the animal was asked to repeat the behaviour twice to ensure that the animal recalled its own behaviour and not the command given for the previous behaviour. All three species could recall their own behaviour significantly better than by chance. The duration for which the animals could recall their behaviour was tested using a staircase paradigm. A delay was implemented between the completion of the behaviour and the command to repeat it. The delay was increased after correct responses and decreased after incorrect responses. The performance of all species fell towards chance level after 12–18 s, with no significant difference between species. These results indicate that sea lions and true seals are aware of their own behaviour and that true seals have similar short-term memory abilities. It also shows that pinnipeds have less developed short-term memory abilities compared to other aquatic predators, such as the bottlenose dolphin. The complexity of pinniped foraging and social behaviour does not seem to have driven the evolution of short-term memory abilities in these animals but might have contributed to their ability to recall their own behaviour.

**Keywords** Cognition · Seal · Sea lion · Self-awareness · Short-term memory

## Introduction

The question if mammals are aware of their own behaviours, and what the importance of such awareness is, has been of interest in several studies (Mercado et al. 1998, 1999; Cowan

2003; Paukner et al. 2007). This question leads to another one, whether mammals possess consciousness, a very intuitively appealing term but difficult to define. Many different definitions of consciousness have been proposed. In an attempt to create some consensus across definitions, Morin (2006) described consciousness not as a clearly defined property, but merely a graded scale of already existing definitions, with the main levels being unawareness (unconsciousness), awareness of the surroundings (consciousness), awareness of inner processes (self-awareness) and aware of being aware of inner processes (meta-self-awareness). On this scale, self-awareness is what most considered to be consciousness. It includes not only perceiving stimuli from the surroundings and responding to them, but also reflections of oneself, both of inner states and of behaviour (Morin 2006).

One method for determining the level of consciousness in an animal is to study its ability to recall its own behaviour. Memory is the requisite for many cognitive abilities. According to Shettleworth (2010), cognition is the ability

---

Kirstin Anderson Hansen and Sara Torres Ortiz contributed equally to this work.

---

✉ Simeon Quirinus Smeele  
simeonqs@hotmail.com

- <sup>1</sup> Department of Biology, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark
- <sup>2</sup> Institute for Terrestrial and Aquatic Wildlife Research (ITAW), University of Veterinary Medicine Hannover, Foundation, Werftstr. 6, 25761 Büsum, Germany
- <sup>3</sup> Dolphin Adventure, Vallarta Adventures, Las Palmas 39, 63732 Nuevo Vallarta, Nayarit, Mexico
- <sup>4</sup> Fjord&Bælt, Margrethes Plads 1, 5300 Kerteminde, Denmark

to receive information, process it, remember it and use it to make decisions. Short-term memory (also including working memory) is used to store information that is processed here and now. Long-term memory is used to store information over a longer period (Baddeley et al. 2010). Previous research has focused on cetaceans, which are believed to have evolved cognitive abilities of the same complexity as primates, in response to the challenging marine environment (Marino 1998, 2002; Boddy et al. 2012). Mercado et al. (1998) showed that two bottlenose dolphins (*Tursiops truncatus*) could repeat their last shown behaviour after a given hand signal (command). The dolphins were capable of repeating four behaviours for which they received training with the repeat command and 32 behaviours for which they were not trained. To make sure that the dolphins were recalling their own behaviour and not the trainer's command, one dolphin was also tested with a double-repeat paradigm, in which she was asked to repeat the just repeated behaviour (sequence being: do a certain behaviour, repeat it, and then repeat it again). The dolphin did this without error. Cowan (2003) continued Mercado's work and showed that the dolphin could remember its own behaviour for 120 s. The only other animal to have been tested in a similar manner is a primate, the macaque (*Macaca nemestrina*), which could remember its own behaviour for delays up to 30 s (Paukner et al. 2007).

Pinnipeds are a large group of marine mammals, which have reinvaded the marine environment. Just like dolphins, pinnipeds are extremely diverse with respect to foraging and social behaviour (Jouventin and Cornet 1980; Bowen et al. 2009). In cetaceans, the complexity of these two factors is positively correlated with brain size (Fox et al. 2017), another often used proxy for cognitive abilities. One may, therefore, wonder if pinnipeds are also able to recall their own behaviour and retain this memory over time.

To investigate this, the short-term memory abilities of several species of pinnipeds were tested with similar experiments to the ones by Mercado et al. (1998) and Cowan (2003). Both sea lions and true seals were tested to investigate if their cognitive abilities were reflecting differences in their social and foraging behaviour.

## Materials and methods

### Animals

The sea lions (SL1–4 hereafter) were kept in Dolphin Adventure in Mexico. They were 13–18 years old and wild-born; SL1–3 were females, SL4 was male. SL3 was in the process of becoming blind and. To avoid the animal being confused by visual cues, this animal was blind-folded for all trials. The grey seal (GS hereafter) was kept at the Marine Biological Research Centre, University of Southern Denmark. GS was 7 years old and born in captivity. The harbour seals (HS1–2 hereafter) were kept in Fjord&Bælt in Denmark. They were 19 and 18 years old and born in captivity; HS1 was male and HS2 female. All animals had been trained for multiple years, but no animal had any previous experience with memory training. They all received their full diet during all training sessions, independent of their performance in the experiments.

### Repeat paradigm

All animals were trained using operant conditioning with positive reinforcement (Pryor et al. 1984). The SLs were trained with a variable-ratio reinforcement schedule, meaning that not all correct responses were primarily reinforced with fish (Kazdin 2012). The seals were trained with a fixed-ratio reinforcement schedule, where all correct behaviours were primarily reinforced with fish (Kazdin 2012). Commands consisted of hand gestures for GS and the HSs, and both hand gestures and verbal commands for each behaviour for the SLs both during training and testing.

The behaviours tested were kept as similar as possible, across species and individuals. There were some differences, as not all animals were trained for the same behaviours (Table 1).

### Baseline

To establish baseline performance for the original behaviours (not the repeat), each animal was given a baseline test. During testing, the assistant would indicate to the trainer

**Table 1** Behaviours used for the repeat paradigms

Behaviour	Description	Animals
Wave	Waving left flipper	All
Snort	Snorting through nose	HS2, GS
Shy	Holding left flipper over eyes	GS, HS1
Chin	Resting chin on deck	HS2
Sing	Growling	SL1, SL2, SL3, SL4, HS1
Spin	Turning around, making a full circle	SL1, SL2, SL3, SL4

what behaviour to ask for by giving the hand gesture, while out of visual range from the animal (an iPad was used for the sea lions, on which the assistant showed the behaviour to be performed). The behaviours were provided from a semi-randomised list, in which each behaviour appeared five times, but never more than twice in a row. The responses of the animal were recorded by the assistant and later checked using video analysis. For an overview of the set-up, see Fig. 1. Whether an animal was ready for the double-repeat paradigm or not did not depend on the performance during the baseline test. This test only served to establish a baseline performance used for comparison with the performance during repeat testing.

### Repeat training

The repeat command consisted of a hand gesture for the GS and HSs, and a hand gesture and verbal command for the SLs. For SL3 only verbal commands were used, since she was trained and tested while blindfolded. The SLs were trained for 3–4 months in 2016. Around 60–80 sessions were used to train the animals to correctly repeat ten behaviours (of which only three were used in the current study). The GS and HSs were trained in the six months before and during testing, and received a total of 100–150 training sessions.

For GS and the HSs, training was divided into four phases, with some overlap between phases. There was a weekly follow-up to keep training as homogeneous as possible across species.

In the first phase, the animal was introduced to the repeat command. The repeat command was a hand signal given with the left hand, whereas the other three behaviour commands were given with the right hand. The repeat command was introduced after the animal had performed one of the three behaviours a couple of times and paired with the behaviour command, given at the same time or just after the repeat command. The repeat command was trained for all three of the original behaviours at the same time to prevent

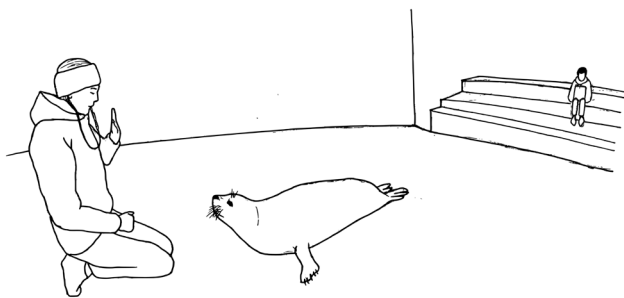
the animal from associating the repeat command with a certain behaviour. In the second phase, the repeat command was given after the repeat behaviour was performed 2–5 times. No original command was presented during or after the repeat command. In the third phase, the three different behaviours were used, intermixed with repeat commands. Each behaviour was only asked for once, after which the repeat command was given. In this phase, the animal was also trained to wait for the next command. The animal was expected to return to the same position after each behaviour and wait for the next command. In the fourth phase, everything was performed as during testing (see below). For training, one out of four training lists with commands was chosen randomly for each session. Four testing lists were used for the four testing sessions.

### Single repeat

Only GS was tested with a single-repeat paradigm. This was done to establish that he understood the meaning of the repeat command. All other animals were tested with a double-repeat paradigm, which also included single-repeat trials. The other animals were not tested with the single-repeat paradigm, since it was clear during training that they had understood the repeat command. The single repeats within the double-repeat paradigm gave enough data to conclude if an animal could correctly repeat the behaviours once.

During testing of the single-repeat paradigm, the same set-up was used as for the baseline testing. The trainer was not aware of the next command on the list and did, therefore, not know if a trial was a repeat or a baseline trial. This was the case in all sessions and all paradigms. The list for the single-repeat trials consisted of 15 trials of two commands. The first command was always one of the three behaviours trained for the repeat paradigm (original behaviours). The second command was a repeat in 50% of the trials (repeat trial) and one of the three original behaviours for the remaining 50% of trials (baseline trial). This was done to prevent the animal from repeating the first command every time regardless of the trial (repeat or baseline). The list was semi-randomised with no behaviour being asked for more than three times in a row (including repeats) and no more than two repeat trials in a row. A total of four sessions were run with a new list each session. In this way, a behaviour was potentially tested ten times with a repeat. However, if the animal responded incorrectly to the first behaviour and the second behaviour was a repeat, this repeat was omitted. If the animal responded incorrectly, no reinforcement was given, and the trainer moved on to the next trial.

Two out of four sessions were made double-blind to demonstrate that the animal was not using accidental cueing by the trainer (e.g. observing the trainer leaning forward when the animal moved towards the correct response). During the



**Fig. 1** Line-drawing of the camera view of the set-up during testing of HS1. Trainer left, assistant right, animal in the middle. The set-ups for data collection with other animals were similar

double-blind session, the trainer wore black glasses (trainer was unable to see the animal's response) while he/she gave the command and the animal responded. A second assistant behind the trainer blew a whistle indicating a correct response or gave a verbal "wrong" if the response was incorrect, since the trainer could not see the animal. The trainer then lifted the glasses to be able to reinforce the animal and see what the next command was, indicated by the first assistant, which stood behind the animal.

### Double repeat

To establish that the animal was remembering its own behaviour and not the last given command, a double-repeat paradigm was used. Here, the animal was asked to repeat an original behaviour twice (a double-repeat trial being: behaviour 1–repeat–repeat). The double-repeat paradigm also contained single-repeat trials (behavior 1–repeat–other behaviour) to establish the performance of single repeats. Additional data for single repeat performance were acquired from the first repeat of a double-repeat trial. There was no specific training for the double-repeat paradigm, since double repeats were already given during the original training. Some sessions were done with lists for training sessions (never the lists used during testing to ensure the animal could not learn the test-sequences) to determine if the animal was performing well.

For the double repeat, the same set-up as for the single repeats (see above) was used. Four sessions were completed with a new list each time. For GS and the HSs, two sessions were double-blind, whereas for the SLs, all sessions were double-blind. The lists consisted of 12 trials with three commands each. 50% of the trials were double repeats. 25% were single repeats (only the second command was a repeat). 25% were non-repeats (all three commands were original behaviours). If the animal responded incorrectly to the first repeat of a double-repeat trial, the second repeat command was not given. The performance of the second repeat was then excluded from the analysis, since it was not asked for, whereas the performance for the first repeat was still included in the analysis.

### Delay

To test the length of time the animals could remember their own behaviour, delays were introduced between the original behaviour and the repeat command. This was done using a staircase paradigm, where the delay increased with 3 s after a correct repeat and decreased with 3 s after an incorrect repeat. The set-up was the same as for the single repeat, with the exception that the list was imported into an R-script (in the statistical programme RStudio, version 1.1.383). This script displayed the to-be-given command for the assistant

after a delay and then asked if the response was correct or not. Based on the answer typed in by the assistant (c for correct, n for not correct), the delay for the next trial was updated and the new command was displayed. Delays were only present before the second command of a trial (both repeats and non-repeats).

### Video analysis

To measure if the animal used certain body postures or behaviours (hereafter called body postures only) to remember how to respond to a repeat command, all sessions were videotaped (each session lasting 3–7 min). A list of all body postures was made. The videos were then checked to determine if a body posture was used for each behaviour throughout a session. Body postures were only recorded between the delivery of reinforcement and the next command, after correct behaviours and not during reinforcement.

For the single-repeat trials, the animal should not expect a repeat command after having repeated the behaviour once. Therefore, the intervals after the repeat commands were not checked for body postures. For the double repeat, only the interval after the first repeat was included in the analysis.

### Statistical analysis

All analyses were done in the statistical programme RStudio (version 1.1.383).

### Single repeat

Repeat performance was the count of correct responses divided by the total number of repeats. Performance was analysed for each behaviour and for the total of all behaviours. To see if performance was above chance, a one-sided binomial test was used (Mercado et al. 1998). The number of correct responses and the total number of repeats were compared to chance performance. Chance performance was set at 1/3 correct, assuming the animal responded with one of the three behaviours tested. The animal could also respond with another behaviour, but this would only decrease the chance performance, which makes 1/3 a conservative estimation. The lower bound of the 95% confidence interval was also calculated.

A generalised linear model with correct and incorrect responses as success/failure and double-blind trials as an explanatory variable was used to compare double-blind sessions with normal sessions.

### Double repeat

For the double-repeat paradigm, performance was analysed for single repeats (including the first repeat of a

double-repeat trial), double repeats and for all repeats grouped together.

To compare species, a generalised linear model with logistic regression was used. Correct and incorrect responses per session were the dependent variable and species was the explanatory variable.

### Delay

To test how long the animal could remember its own behaviour, a staircase paradigm was used with delays between the original behaviour and the repeat command. The set-up was the same as in the single repeat, with 15 trials, nine with delays and six without. The behaviours were read from a computer by the assistant to the trainer. An R-script displayed the behaviour after the correct delay and asked if the response was correct or not. The delay was adjusted, depending on the answer given by the assistant. Delays were present before behaviour two of each trial, regardless of it being a repeat or not. The staircase started with no delay, increased 3 s after a correct response and decreased 3 s after an incorrect response. Because of the time necessary to reinforce the animal, record the response and communicate the next trial, the actual delay experienced by the animal was longer than the delay in the script. The actual length of the delay was measured from the video recordings. The delay started when the trainer blew the whistle indicating a correct response after the first behaviour and ended when the trainer gave the command for the repeat.

No training was done prior to introducing delays between the original behaviour and the repeat command. All phocids showed some form of frustration. GS became agitated during the first delay session. To prevent frustration, the number of delay trials (normally all trials had a delay) was build up throughout many sessions, starting with one delay trial for every two non-delay trials. HS1 and HS2 were given delays on all trials for all sessions. HS1 did, however, show poor performance on training sessions without delays, the day after a session with delays. HS2 showed frustration during data collecting (jaw-clapping) and began to anticipate repeat commands during the delay. She would, e.g. offer the behaviour *wave* during the delay. Trials where she anticipated were removed from analysis.

### Video analysis

To analyse if body postures had a positive influence on performance, a list was made with the count of how many times each body posture was observed after each behaviour. For the double repeat, the first repeat was considered as the behaviour that should be repeated. For each body posture, the behaviour after which it was used most often was found. If it was used equally often after more than one behaviour, it

was not considered statistically, since it would not be helpful as a memory tactic. When a behaviour was found, the performance was analysed on the repeat trials following this behaviour with and without the body posture during inter-trial periods. These two groups were compared using a one-sided binomial test, where correct responses and total number of trials with the body posture were compared to performance without the body posture as hypothesised probability of success.

For the delay paradigm, only the delay interval was analysed.

## Results

### Single repeat

The single-repeat paradigm was only completed with GS (see Table 2). Due to time constraints, no single-repeat paradigm was done with HSs and SLs; however, they received single repeats during the double-repeat paradigm. GS performed above chance (one-sided binomial test,  $p < 0.05$ ) for snort, wave and the total of all behaviours. The difference in performance between non-blind and double-blind trials was not significant (one-sided binomial test,  $p > 0.1$ ).

### Double repeat

For double repeats, all animals showed overall performance above chance level (33% correct, see Table 3). SL1 performed around chance for the spin and did wave instead of repeating the spin. GS performed below chance on wave (23% correct). The other animals performed above chance for all behaviours when looking at the total number of double repeats.

All animals but SL2 and SL3 scored best on the double repeats (see Fig. 2). The HSs had the best overall performance, followed by the SLs. The poorer performance by

**Table 2** Performance (measured as the amount of correct out of the total amount of trials) for GS during repeat trials

Behavior	Performance	Incorrect responses		
		Shy	Snort	Wave
Shy	5/9 (0.25)	0	2	1
Snort	8/10* (0.49)	1	0	0
Wave	7/10* (0.39)	2	1	0
Total	20/29* (0.52)	3	3	1

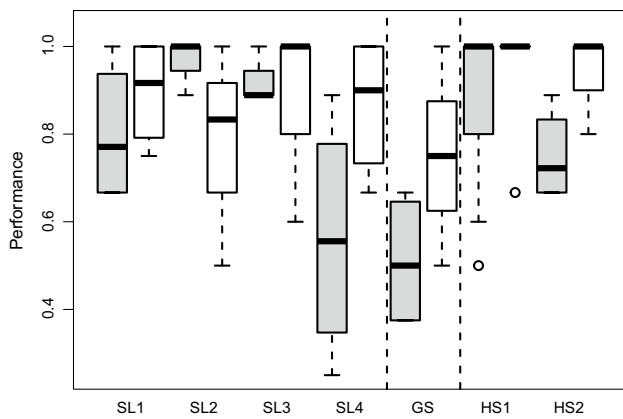
An asterisk indicates performance significant above chance (one-sided binomial test,  $p < 0.05$ ). Lower bound of 95% confidence interval is given in parentheses. Last three columns show number of each incorrect baseline behaviour during the repeat trials. Incorrect responses other than the three baseline behaviours are not shown

**Table 3** Performance (measured as the amount of correct out of the total amount of trials) for single (from single-repeat trials and the first repeat of a double-repeat trial), double (the second repeat of a double-repeat trial) and the total (the combination of the two) of the repeat trials for all animals

SL1				Incorrect responses		
Behaviour	Single	Double	Total	Sing	Spin	Wave
Sing	12/12* (0.78)	7/8* (0.53)	19/20* (0.78)			1
Spin	4/11 (0.14)	2/3 (0.14)	6/14 (0.19)			8
Wave	12/12* (0.78)	7/7* (0.65)	19/19* (0.85)			
Total	28/35* (0.66)	16/18* (0.72)	44/53* (0.72)			9
SL2				Incorrect responses		
Behaviour	Single	Double	Total	Sing	Spin	Wave
Sing	11/12* (0.66)	8/8* (0.69)	19/20* (0.78)			1
Spin	12/12* (0.78)	5/8 (0.29)	17/20* (0.66)			1
Wave	12/12* (0.78)	6/8* (0.40)	18/20* (0.72)		1	
Total	35/36* (0.87)	19/24* (0.61)	54/60* (0.81)		1	2
SL3				Incorrect responses		
Behaviour	Single	Double	Total	Sing	Spin	Wave
Sing	12/12* (0.78)	8/8* (0.69)	20/20* (0.86)			
Spin	11/12* (0.66)	7/8* (0.53)	18/20* (0.72)			2
Wave	10/12* (0.56)	6/7* (0.48)	16/19* (0.64)	1	1	
Total	33/36* (0.8)	21/23* (0.75)	54/59* (0.83)	1	1	2
SL4				Incorrect responses		
Behaviour	Single	Double	Total	Sing	Spin	Wave
Sing	9/12* (0.47)	5/6* (0.42)	14/18* (0.56)			3
Spin	6/11 (0.27)	3/4 (0.25)	9/15* (0.36)	5		1
Wave	5/12 (0.18)	4/4* (0.47)	9/16* (0.33)	5	1	
Total	20/35* (0.42)	12/14* (0.61)	32/49* (0.53)	10	1	4
GS				Incorrect responses		
Behaviour	Single	Double	Total	Snort	Shy	Wave
Snort	8/11* (0.44)	5/5* (0.55)	13/16* (0.58)		3	
Shy	6/11 (0.27)	4/5* (0.34)	10/16* (0.39)	5		
Wave	3/11 (0.08)	0/2 (0)	3/13 (0.07)	3	5	
Total	17/33* (0.36)	9/12* (0.47)	26/45* (0.38)	8	8	
HS1				Incorrect responses		
Behaviour	Single	Double	Total	Sing	Shy	Wave
Sing	9/12* (0.47)	6/6* (0.61)	15/18* (0.62)			3
Shy	10/11* (0.64)	5/6* (0.42)	15/17* (0.67)	1		1
Wave	12/12* (0.78)	8/8* (0.69)	20/20* (0.86)			
Total	31/35* (0.76)	19/20* (0.78)	50/55* (0.82)	1		4
HS2				Incorrect responses		
Behaviour	Single	Double	Total	Snort	Chin	Wave
Snort	7/12 (0.32)	4/5* (0.34)	11/17* (0.42)		1	4
Chin	11/12* (0.66)	7/7* (0.65)	18/19* (0.77)	1		
Wave	9/12* (0.47)	7/7* (0.65)	16/19* (0.64)		2	
Total	27/36* (0.60)	18/19* (0.77)	45/55* (0.71)	1	3	4

An asterisk indicates performance significant above chance (one-sided binomial test,  $p < 0.05$ ). Lower bound of 95% confidence interval in parentheses. Last three columns show incorrect behaviour during the repeat trials. Incorrect responses other than the three baseline behaviours are not shown





**Fig. 2** Boxplots of performance (measured as the amount of correct out of the total amount of trials) per session for all animals for single (filled boxes) and double (not filled boxes) repeats

GS compared to the HSs and SLs is significant (generalised linear model with logistic regression,  $p < 0.01$ ). The difference between HSs and SLs is not significant (generalised linear model with logistic regression,  $p > 0.1$ ).

For the seals, there was no difference between non-double-blind and double-blind sessions (generalised linear model with logistic regression,  $p > 0.1$ ). For sea lions, all sessions were double-blind.

**Delays**

The HSs together with SL4 scored best with performance significant above chance level (one-sided binomial test,  $p < 0.05$ ) for delays up to nine seconds (see Table 4). GS

together with SL1 and SL3 performed significant above chance level for delays up to 6 s. SL2 only performed just above chance level for delays of 3 s.

When correct and incorrect responses are modelled with a generalised linear model with logistic regression, SL2 and SL3 clearly show lower performance than the other animals (Fig. 3). The modelled performance crossed chance-level performance between 4 and 7 s for SL2 and SL3, and between 12 and 18 s for SL1, SL4, GS and HS1. For HS2, not enough data points were available to model the performance until it crossed chance level.

**Video analysis**

No body posture had positive influence on performance during the double-repeat paradigm (one-sided binomial test,  $p > 0.05$ ).

During the delay paradigm, only for GS did a body posture have positive influence on performance. GS shifted his weight to the left, often rolling his left flipper under his body, after snorts. This had a positive influence on his performance (one-sided binomial test,  $p < 0.05$ ). On trials where he shifted his weight, he got all the 14 repeat trials correct. Without the weight shift, he got four out of eight repeat trials correct.

**Discussion**

The results clearly show that all three species of tested pinnipeds can recall their own behaviour.

**Table 4** Performance (measured as the amount of correct out of the total amount of trials) per delay (rounded to the nearest multiple of three) for all animals

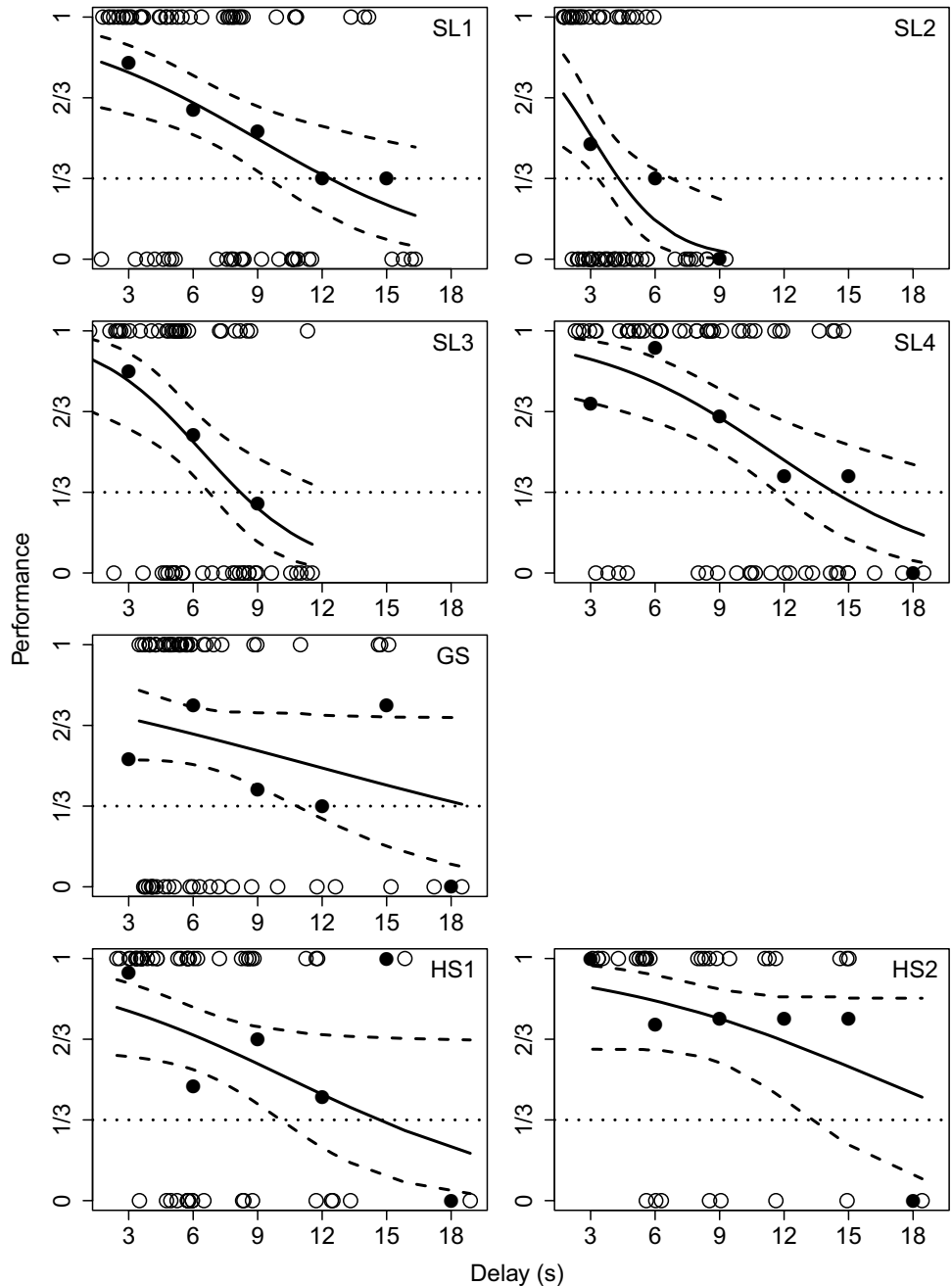
Delay (s)	SL1	SL2	SL3	SL4
3	17/21* (0.62)	20/42* (0.34)	11/13* (0.59)	7/10* (0.39)
6	8/13* (0.35)	5/15 (0.14)	16/28* (0.40)	13/14* (0.70)
9	10/19 (0.32)	0/6 (0.14)	4/14 (0.10)	11/17* (0.42)
12	4/12 (0.12)			4/10 (0.15)
15	2/6 (0.06)			4/10 (0.15)
18				0/2 (0.15)
Delay (s)	GS	HS1	HS2	
3	10/19 (0.32)	16/17* (0.75)	5/5* (0.55)	
6	24/32* (0.59)	9/19 (0.27)	8/11* (0.44)	
9	2/5 (0.08)	6/9* (0.34)	6/8* (0.40)	
12	1/3 (0.02)	3/7 (0.13)	3/4 (0.25)	
15	3/4 (0.25)	1/1 (0.05)	3/4 (0.25)	
18	0/2 (0)	0/1 (0)	0/1 (0)	

An asterisk indicates performance significant above chance (one-sided binomial test,  $p < 0.05$ ). Lower bound of 95% confidence interval in parentheses

Only two animals had average performance below 80% during double-repeat trials (see Fig. 2). SL4 performed very well on the first two sessions but performed poorly on the last two sessions. There are two reasons for this decreased performance. SL4 was under eye treatment and was tested in a confined space in the last two sessions, where he refused to do spins on multiple trials. For GS, multiple factors contributed to poor performance. In general, his performance was much more variable, which is also visible in his baseline performance. Nevertheless, his overall performance is still well above chance level, thereby fully supporting our conclusions. Here, GS did

85% of its behaviours correct, where all other animals did above 95% of their behaviours correct. Compared to the other animals, GS was younger and had only 4 years of training experience. Both the GS and the HSs were trained with a larger reinforcement after correct performance on repeat trials, compared to original behaviours. For GS, this difference was largest (three/four pieces of herring vs one sprat). Double-repeat trials were, therefore, very reinforcing for one behaviour (one sprat/capelin for the original behaviour, three pieces of herring for the first repeat and four pieces of herring for the second repeat), which might explain why GS often stuck to performing the behaviour

**Fig. 3** Performance (0 is incorrect, 1 is correct) of all trials during delay paradigm (empty circles). Solid lines are fitted values from a generalised linear model with logistic regression. Dashed lines are upper and lower bounds of the 95% confidence interval. Full circles are average performance (measured as the amount of correct out of the total amount of trials) on delays rounded to the nearest multiple of three. Dotted line is chance performance (1/3 correct)





from the first double repeat he encountered on the following few trials. This difference could also explain why his performance on the single repeat fell drastically during the double-repeat paradigm (69% repeat trials correct during the single-repeat paradigm; 51% single-repeat trials correct during the double-repeat paradigm). One concern could be that GS relied on reference memory rather than episodic-like memory during the double-repeat paradigm. In other words, he could get set on a certain behaviour after a double repeat and remember that performing this behaviour gave reinforcement after a repeat command. To test if this was the case, his actual responses were compared to modelled responses where either episodic-like memory was used (always repeating the last performed behaviour in a repeat trial) or reference memory was used (always repeating the behaviour from the last double-repeat trial). The episodic-like memory model predicted the actual response best (52% matching vs 45% matching for reference memory model). This makes it highly likely that, even though GS might have been confused due to the high reinforcement value of a double-repeat trial, he was still using episodic-like memory to solve the paradigm.

Mercado et al. (1998) found similar results for two bottlenose dolphins with respect to the single repeat of four behaviours (87% and 62% correct). Additionally, animals were tested for 32 behaviours that were not previously trained with the repeat paradigm. Here, animals scored similarly (90% and 57%). For the double-repeat paradigm, only the best scoring animal was tested for the four trained behaviours. This animal scored 100% on this paradigm. Performance of the bottlenose dolphins is hard to compare with that of the pinnipeds, because no training effort was reported. However, the performance of the pinnipeds matches that of the bottlenose dolphins very well, with the total performance of GS at 58% as the lowest and that of HS1 at 91% as the highest (see Table 3).

During the delay trials, GS tended to shift his weight to the left (supporting his upper body on his left flipper, leaving the right flipper free to wave or do shy) after snorts. He performed better on trials where he shifted his weight than on trials where he did not. Whether he did this consciously to remember he should snort on the next trial, or just because he did not expect a wave or shy (for which he had to use his left flipper), or found the position to be more comfortable, is not clear. To make results significant for GS, the delay paradigm will have to be repeated and he should be trained to stay in the same position during the delay. Cowan (2003) had similar issues during longer delays with bottlenose dolphins. To try to remove the possibility to use body posture as a clue, he did a second experiment, with a 30-s delay, where the dolphin was asked to press a paddle at either three or 14 s after the first behaviour was completed. The performance was still

above chance, but degraded quite a bit (29% correct, vs 93% correct without paddle pressing with the same delay).

Due to time constraints, only a limited number of delay paradigm sessions were made. The modelled performance showed at which delay the animals were expected to perform at chance level. Due to the few trials, the 95% confidence intervals were very broad at longer delays for the phocids, which makes the estimation of maximum delays difficult (see Fig. 2). The fact that the lower bound of the 95% confidence interval shows up to a 10-s difference where it crosses chance level is caused by too few replicates. Still, our results show that performance declines towards chance-level performance after 15–18 s, which matches the performance of another harbour seal tested with delayed-matching-to-sample (DMTS) (Mauck and Dehnhardt 2007). When tested in a regular DMTS paradigm, where the animal had to match one out of two comparison stimuli to a sample stimulus after a delay, the performance declined towards chance-level performance after 12 s. This decline was not apparent when also spatial information was available (where the stimulus was placed in the enclosure). These results indicate that harbour seals have a much better developed spatial memory, than general memory. They may need this well-developed memory to navigate over long distances and find, e.g. the exact beach where they left their young.

The SLs performance was more varied. SL2 was being trained for another paradigm at the time the delay paradigm was tested. Since she also had low performance on the short delays, it is likely that she did not remember the repeat paradigm very well. SL3 was tested while blindfolded, since she was in the process of becoming blind. This provides an even stronger proof that she was not using accidental cueing but may also have impaired her performance during the delay paradigm, since she had to rely on her hearing and touch senses to know what was happening around her. This likely consumed more attention and therefore impaired performance. The other two sea lions performed as well as the harbour seals. The 95% confidence interval is much narrower for the sea lions than for the HSs and GS and shows that the 12–18 s interval is very likely the retention time for own behaviour with the current set-up.

The performance half-time of sea lions using the DMTS paradigm is well above 30 s (Pack et al. 1991). DMTS performance is often compared using zero-delay performance and performance half-time. For the data of this study, performance half-time is very close to chance performance for GS, but much higher for the HSs. This is due to the HSs modelled zero-delay performance being around 90%. The memory of GS (measured as performance half-time) is better than that of the HSs, but motivation and concentration resulted in lower overall performance. However, the large

difference between the performance of HS1 and HS2 suggests that more data is required to draw solid conclusions on this matter. A reason for the much higher performance half-time for the sea lions than the phocids could be the use of an inter-trial interval in the sea lion testing. Pack et al. (1991) waited 30 s between trials. Such long inter-trial intervals have been shown to increase performance, by decreasing interference from memory from previous trials (Zentall and Smith 2016).

The only other studies investigating short-term memory by recall of own behaviour after a delay found that bottlenose dolphins performed above chance level even after 2 mins (Cowan 2003), and macaques after 30 s (Paukner et al. 2007). Apart from the fact that dolphins and primates have been shown to have well-developed cognitive abilities (Marino 2002), and therefore might have better memory of their own behaviour, there are two factors that could account for the much longer retention time by these species. The bottlenose dolphins had been trained for much longer on the repeat paradigm and multiple tests had been done before the delay paradigm was tested. A second factor is the inter-trial interval. The dolphin had to wait at least 30 s (the macaques 20 s) between trials and 2 mins between every four trials. This has a positive influence on retention time (Zentall and Smith 2016) and might very well explain the difference in performance.

The second aim was to compare otariids with phocids and two closely related phocids with each other. Otariids normally breed on islands and have harems between five and 15 females. The mothers stay with the pup for up to a year. For some species, cooperation during hunting has been recorded. Phocids breed on different substrates and show a large variety of breeding strategies. Weaning of the pup occurs within a few of weeks for most species (Jouventin and Cornet 1980). No cooperation during hunting has been reported. This general trend is well represented by the species in the current study (all three studied species have life-histories similar to the general life-history of the family). Large South American sea lion males defend multiple females during breeding. Smaller individuals may develop alternative strategies to be able to breed. Females lactate for up to 10 months, sometimes staying with their young after lactation ceases (Cappozzo and Perrin 2009). Hunting in groups has been observed (Sepúlveda et al. 2007). The phocid species used in this study are either polygamous (most grey seal populations) or solitary (Baltic grey seals and harbour seals). They lactate their young for only a few weeks, then leaving them to forage for themselves. No hunting in groups has been observed (Burns 2009; Hall and Thompson 2009). These differences in ecology would make it likely for otariids to have evolved better cognitive abilities, since social complexity and especially pair-bonding have been reported by Dunbar

and Shultz (2007) as a driver for larger brains (a proxy used as a measure of cognitive abilities).

There is no clear difference between the otariids and phocids, neither in the double repeat nor the delay paradigm. Another otariid, the Californian sea lion (*Zalophus californianus*) has a short-term memory duration well beyond 30 s (Pack et al. 1991), which is much higher than what is found in the harbour seal (Mauck and Dehnhardt 2007). It is likely that phocids do not have the same short-term memory duration as otariids, since the latter might have developed a better general short-term memory, due to the challenges posed by their foraging and social behaviour. It could, e.g. be that females would benefit from increased short-term memory in developing a strategy to handle their offspring or that otariids hunting together would benefit from being able to remember which moves they made. Our study showed no difference in short-term memory for own behaviour is in contrast with these previous findings. It might be that the difference is caused by differences in training between the otariids and phocids. However, this is unlikely since most animals did very well on the short delays, making it likely that they were all trained very well for the repeat paradigm. Our study is using exactly the same methodology for both otariids and phocids, which was not the case in previous studies. Therefore, the more likely explanation is that the difference in retention in previous studies with DMTS is caused by differences in the methods of those studies; and that the short-term memory of otariids and phocids is actually very similar. The more complex mating system of otariids does not seem to have led to improved memory capabilities. Instead, the main reason for developing short-term memory abilities may be found in the foraging behaviour of seals, which may be quite similar across otariids and phocids.

As for the phocids, one would expect the grey seal to have a better developed memory for own behaviour than harbour seals. In spite of being closely related, they are quite different with respect to social life and foraging strategies (Jouventin and Cornet 1980; Higdon et al. 2007). Harbour seals are more conservative in their foraging, mainly going after few fish species and cephalopods (Hall et al. 1998; Berg et al. 2002). Grey seals also eat mainly fish and cephalopods (Mikkelsen et al. 2002; Lundström et al. 2007), but especially males are keen on raiding salmon traps (Lehtonen and Suuronen 2010; Königson et al. 2013) and have been observed to take harbour seals (Bishop et al. 2016; Brownlow et al. 2016), harbour porpoises (Haelters et al. 2012; Leopold et al. 2015; Heers et al. 2018) and even young grey seals (van Neer et al. 2015). Socially, grey seals are much more diverse than harbour seals. They can both be seemingly monogamous (in the Baltic Sea) and polygamous (in the Atlantic). How the females are defended by the males depends on the substrate (Jouventin and Cornet 1980). To attract females, harbour seals mainly produce

roars and otherwise use their flippers to produce sounds. Grey seals produce at least seven distinct vocalisations and appear much more flexible in their vocalisations (Asselin et al. 1993). Both species are capable of vocal mimicry, even though there is only evidence from one harbour seal (Ralls et al. 1985; Stansbury 2015). Although they are only present in the North-Atlantic, North Sea and Baltic Sea, grey seals appear to be the more opportunistic species. Therefore, they would benefit most with increased cognitive abilities. This adaptive life-style would benefit greatly from good memory for own behaviour. In another study comparing performance on the mirror test (Martín 2016), greys seals showed more self-directed behaviour in front of a mirror, also indicating more self-awareness. In the current study, one harbour seal (HS2) seemed to remember her behaviour better than HS1 and the GS (see Fig. 2). This is not what was expected, given the expected higher cognitive skills of grey seals compared to harbour seals, from their social and foraging behaviour. The reason for this may be that GSs would be better off than HSs in other cognitive abilities than the ones measured here. Also, cognitive tests may be affected by the animal's ability to be patient and voluntarily participating in many trials, and there may be innate differences between the two species in this respect, as well as differences in their training and husbandry history that may affect the results, besides differences in their cognitive abilities. Also, as only one GS was tested, there is also the possibility that this individual is not representative in cognitive and memory abilities with grey seals in general.

There is clear evidence that the increased cognitive abilities and brain size of cetaceans, at least partly, are due to evolutionary pressure caused by their complex social structures (Fox et al. 2017). Their cognitive abilities may also have led to more diverse foraging strategies. Marino (2002) argues that increased social complexity could have been a response to predation during development of the young. Pinnipeds can prevent predation from the marine environment by pupping on land, even though this leaves them with some terrestrial predators. However, the offspring need to return to the marine environment at some point. This overlap in ecology with cetaceans would make it likely for pinnipeds to evolve more complex social structures and therefore develop complex cognitive abilities. Pinnipeds face the same challenges as cetaceans to find prey in a three-dimensional space. Besides cetaceans, this has presumably also led to increased brain size in chiropterans and primates (Eisenberg and Wilson 1978; Clutton-Brock and Harvey 1980). If the evolutionary pressure has been the same for pinnipeds as it was for cetaceans since their return to the marine environment, there ought to be some sign of this. The ability to recall own behaviour has not been tested for other animals than primates and cetaceans, even though some similar studies have been done with rats and pigeons (Beninger et al. 1974;

Shimp 1982). Without more data from other mammals, it is difficult to understand the significance and evolutionary driving forces of this ability. This ability does, however, require some form of self-awareness, since the response is based on the animal's own behaviour.

## Conclusion

All pinniped species tested could recall their own behaviour. This indicates self-awareness with respect to own behaviour in both otariids and phocids. In the current set-up, the retention time for own behaviour of the pinnipeds tested was 12–18 s (with the exception of SL2 and SL3). There was no difference between otariids and phocids, which stands in contrast with findings of previous studies using DMTS. Since our study was conducted with very similar methods for all species, it is likely that there is no difference in short-term memory between otariids and phocids; and that previous differences are caused by differences in the methods of those studies. The retention of 12–18 s is much shorter than what is found in macaques and bottlenose dolphins, probably due to methodological differences, such as longer training before testing and longer inter-trial intervals. The complexity of pinniped foraging and social behaviour does not seem to have driven the development of short-term memory abilities in these animals.

**Acknowledgements** We thank Wayne Phillips, Benjamin Doshner and Omar del Oso from Dolphin Adventure for their support and for the pre-disposition of the animals in Dolphin Adventure. We also thank Abel Reyes, Jesus Manuel Perez, Jesus Silva, Kelli Rust, Luis Hernández and Paola Valle for training the animals and providing their session time with them.

**Funding** This study was funded by the Danish Research Council (Grant No. 400200536) and the Institute for Terrestrial and Aquatic Wildlife Research (ITAW) University of Veterinary Medicine Hannover, Foundation (Grant No. 60079084).

## Compliance with ethical standards

**Conflict of interest** All authors declare that they have no conflict of interest.

**Ethical approval** All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. The animals from Dolphin Adventure were kept under permits from Mexican Secretariat of Environment and Natural Resources [permit: INE/CITES/DGVS-EF-P-0033-NAY/00 (PIMVS)]. The animals from Fjord&Bælt (permit: J.nr.SVANA-610-00084) and The University of Southern Denmark (permit: 2300-50120-00003-09) were kept under permits from the Danish Ministry of Food and Agriculture. The experiments did not require an application to the Animal Ethics Committee of neither Denmark nor Mexico, as animals participated voluntarily in the experiments and were not affected by them in any way. This article does not contain any studies with human participants performed by any of the authors.

## References

- Asselin S, Hammill MO, Barrette C (1993) Underwater vocalizations of ice breeding gray seals. *Can J Zool* 71(11):2211–2219
- Baddeley A, Eysenck MW, Anderson MC (2010) *Memory*. East Sussex, Psychology Press, Hove
- Beninger RJ, Kendall SB, Vanderwolf CH (1974) The ability of rats to discriminate their own behaviours. *Can J Psychol/Rev Can Psychol* 28(1):79–91
- Berg I, Haug T, Nilssen KT (2002) Harbour Seal (*Phoca vitulina*) Diet in Vester » len, North Norway. *Sarsia N Atl Mar Sci* 87(6):451–461
- Bishop AM, Onoufriou J, Moss S, Pomeroy PP, Twiss SD (2016) Cannibalism by a Male Grey Seal (*Halichoerus grypus*) in the North Sea. *Aquat Mamm* 42(2):137–143
- Boddy A, McGowen M, Sherwood C, Grossman L, Goodman M, Wildman D (2012) Comparative analysis of encephalization in mammals reveals relaxed constraints on anthropoid primate and cetacean brain scaling. *J Evol Biol* 25(5):981–994
- Bowen WD, Beck CA, Austin DA (2009) Pinniped ecology. In: Perrin WF, Würsig B, Thewissen JGM (eds) *Encyclopedia of marine mammals*. Elsevier, San Diego, pp 533–541
- Brownlow A, Onoufriou J, Bishop A, Davison N, Thompson D (2016) Corkscrew seals: grey seal (*Halichoerus grypus*) infanticide and cannibalism may indicate the cause of spiral lacerations in seals. *Plos One* 11(6):e0156464
- Burns JJ (2009) Harbor Seal and spotted seal *Phoca vitulina* and *P. largha*. In: Perrin WF, Würsig B, Thewissen JGM (eds) *Encyclopedia of marine mammals*. Elsevier, San Diego, pp 533–541
- Cappozzo HL, Perrin WF (2009) South American Sea Lion: *Otaria flavescens*. In: Perrin WF, Würsig B, Thewissen JGM (eds) *Encyclopedia of marine mammals*, 2nd edn. Academic Press, London, pp 1076–1079
- Clutton-Brock TH, Harvey PH (1980) Primates, brains and ecology. *J Zool* 190(Mar):309–323
- Cowan RB (2003) Short term memory for behavior in the Bottlenose Dolphin (*Tursiops truncatus*). *Psychology*. University of Hawaii at Manoa, Manoa, p 80
- Dunbar RI, Shultz S (2007) Evolution in the social brain. *Science* 317(5843):1344–1347
- Eisenberg JF, Wilson DE (1978) Relative brain size and feeding strategies in the Chiroptera. *Evolution* 32(4):740–751
- Fox KC, Muthukrishna M, Shultz S (2017) The social and cultural roots of whale and dolphin brains. *Nat Ecol Evol* 1(11):1699
- Haelters J, Kerckhof F, Jauniaux T, Degraer S (2012) The Grey Seal (*Halichoerus grypus*) as a Predator of Harbour Porpoises (*Phocoena phocoena*)? *Aquat Mamm* 38(4):343–353
- Hall A, Thompson D (2009) Grey seal *Halichoerus grypus*. In: Perrin WF, Würsig B, Thewissen JGM (eds) *Encyclopedia of marine mammals*. Elsevier, San Diego, pp 501–503
- Hall AJ, Watkins J, Hammond PS (1998) Seasonal variation in the diet of harbour seals in the south-western North Sea. *Mar Ecol Prog Ser* 170:269–281
- Heers T, van Neer A, Becker A, Gross S, Hansen KA, Siebert U, Abdulmawjood A (2018) Loop-mediated isothermal amplification (LAMP) as a confirmatory and rapid DNA detection method for grey seal (*Halichoerus grypus*) predation on harbour porpoises (*Phocoena phocoena*). *J Sea Res* 140:32–39
- Higdon JW, Bininda-Emonds OR, Beck RM, Ferguson SH (2007) Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol Biol* 7(1):216
- Jouventin P, Cornet A (1980) The sociobiology of pinnipeds. *Adv Study Behav* 11:121–141
- Kazdin AE (2012) Behavior modification in applied settings. Waveland Press, Long Grove
- Konigson S, Fjalling A, Berglind M, Lunneryd SG (2013) Male gray seals specialize in raiding salmon traps. *Fish Res* 148:117–123
- Lehtonen E, Suuronen P (2010) Live-capture of grey seals in a modified salmon trap. *Fish Res* 102(1–2):214–216
- Leopold MF, Begeman L, van Bleijswijk JDL, Ijsseldijk LL, Witte HJ, Grone A (2015) Exposing the grey seal as a major predator of harbour porpoises. *Proc R Soc B Biol Sci* 282(1798):20142429
- Lundström K, Hjerne O, Alexandersson K, Karlsson O (2007) Estimation of grey seal (*Halichoerus grypus*) diet composition in the Baltic Sea. *NAMMCO Sci Publ* 6:177–196
- Marino L (1998) A comparison of encephalization between odontocete cetaceans and anthropoid primates. *Brain Behav Evol* 51(4):230–238
- Marino L (2002) Convergence of complex cognitive abilities in cetaceans and primates. *Brain Behav Evol* 59(1–2):21–32
- Martín HF (2016) Self-cognition in marine mammals studied by the mirror test. Master thesis, Department of Biology, Odense University of Southern Denmark
- Mauck B, Dehnhardt G (2007) Spatial multiple-choice matching in a harbour seal (*Phoca vitulina*): differential encoding of landscape versus local feature information? *Anim Cogn* 10(4):397–405
- Mercado E, Murray SO, Uyeyama RK, Pack AA, Herman LM (1998) Memory for recent actions in the bottlenosed dolphin (*Tursiops truncatus*): repetition of arbitrary behaviors using an abstract rule. *Anim Learn Behav* 26(2):210–218
- Mercado E, Uyeyama RK, Pack AA, Herman LM (1999) Memory for action events in the bottlenosed dolphin. *Anim Cogn* 2:17–25
- Mikkelsen B, Haug T, Nilssen KT (2002) Summer diet of grey seals (*Halichoerus grypus*) in Faroese waters. *Sarsia N Atlant Mar Sci* 87(6):462–471
- Morin A (2006) Levels of consciousness and self-awareness: a comparison and integration of various neurocognitive views. *Conscious Cogn* 15(2):358–371
- Pack AA, Herman LM, Roitblat HL (1991) Generalization of visual matching and delayed matching by a California sea lion (*Zalophus californianus*). *Learn Behav* 19(1):37–48
- Paukner A, Anderson JR, Donaldson DI, Ferrari PF (2007) Cued repetition of self-directed behaviors in macaques (*Macaca nemestrina*). *J Exp Psychol Anim Behav Process* 33(2):139
- Pryor K, Callenbach EBE, Davis GW (2002) Don't shoot the dog: how to improve yourself and others through behavioral training, revised edition. Ringpress Books Ltd., Gloucestershire
- Ralls K, Fiorelli P, Gish S (1985) Vocalizations and vocal mimicry in captive harbor seals, *phoca-vitulina*. *Can J Zool* 63(5):1050–1056
- Sepúlveda M, Pérez M, Sielfeld W, Oliva D, Durán L, Rodríguez L, Araos V, Buscaglia M (2007) Operational interaction between South American sea lions *Otaria flavescens* and artisanal (small-scale) fishing in Chile: results from interview surveys and on-board observations. *Fish Res* 83(2–3):332–340
- Shettleworth SJ (2010) *Cognition, evolution, and behavior*. Oxford University Press, Oxford
- Shimp CP (1982) On metaknowledge in the pigeon: an organism's knowledge about its own behavior. *Anim Learn Behav* 10(3):358–364
- Stansbury A (2015) *Vocal learning and development in the grey seal, Halichoerus grypus*. University of St Andrews, St Andrews
- van Neer A, Jensen LF, Siebert U (2015) Grey seal (*Halichoerus grypus*) predation on harbour seals (*Phoca vitulina*) on the island of Helgoland, Germany. *J Sea Res* 97:1–4
- Zentall TR, Smith AP (2016) Delayed matching-to-sample: a tool to assess memory and other cognitive processes in pigeons. *Behav Proc* 123:26–42

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.