Summer 7-9-2018

An Analysis of Innovate Training with Bottlenose Dolphins (Tursiops truncatus)

Raymond John Van Steyn
CUNY Hunter College

An Analysis of Innovate Training with Bottlenose Dolphins (*Tursiops truncatus*)

by

Raymond John Van Steyn

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Animal Behavior and Conservation, Hunter College
The City University of New York

2018

	Thesis Sponsor:
8.6.2018	Diana Reiss, Ph.D.
Date	Signature
8.6.2018	Martin Chodorow, Ph.D.
Date	Signature of Second Reader

Dedication

For Mom and Dad. Thank you for never giving up after all the parent-teacher conferences.

For Alanna. Thank you for all the love and support throughout it all.

Acknowledgements

There are a host of people who were integral to making this project a reality. Firstly, I would like to thank my mentor, Dr. Diana Reiss for her patience, guidance and support during this thesis and my training. I would also like to thank my second reader, Dr. Martin Chodorow whose counsel throughout this project was invaluable. I would like to thank everyone in Diana Reiss' Cetacean Cognition and Communication laboratory for their support and guidance, particularly Adrienne Koepke and Jennifer Savoie. I would like to thank my friends, mentors and colleagues at Rise for their support and guidance. I would also like to thank the training staff at the National Aquarium for their support and hospitality throughout this project. Lastly, I would like to thank all of my friends, colleagues and professors in the Hunter College Animal Behavior and Conservation program. It has truly been an honor to learn from you all, and to walk these halls with you. Thank you.

Abstract

The National Aquarium in Baltimore, Maryland conducted a training program in 2014 to develop a gestural command for their dolphins called “innovate”. This training paradigm was developed to resemble the seminal research by Pryor, Haag and O’Reilly (1969), as well as more recent efforts of Braslau-Schneck (1993) and Kuczaj and Eskelinen (2014) of training dolphins to offer “creative” behaviors not developed through conventional methods of behavioral modification, such as shaping. The goal of the present study was to observe records taken during the National Aquarium’s training procedure as well as data collected ~3 years after said training in order to analyze and observe training practices and resulting learning of the task among the trainers and dolphins. All subjects developed an understanding of the task “do something different” and exhibited patterns in the learning process similar to those reported by Pryor, Haag and O’Reilly (1969), as well as strategies used to complete this task similar to those reported by Kuczaj and Eskelinen (2014).

Keywords: innovate, bottlenose dolphin, Tursiops truncatus, learning, behavior, training, memory

Table of Contents

List of Figures

List of Tables

An Analysis of Innovate Training with Bottlenose Dolphins (*Tursiops truncatus*)

Behavioral demonstrations for the public are a common part of the daily life of most captive dolphins. Most behaviors exhibited in these demonstrations are either a part of the natural repertoire of dolphins or those exhibited by their wild counterparts but in either case, the animals are trained to perform a specific behavior when given a specific signal (cue) by a trainer. Most often training is done through shaping (Skinner 1951) or capturing a complete behavior and delivering a reinforcer (Pryor, Haag & O'Reilly 1969). When this cue is given to the dolphin, they will perform the corresponding behavior for food or another type of reinforcement. In this training paradigm, each cue is associated with one specific behavior.

In a seminal study, Pryor, Haag & O'Reilly (1969) sought to increase behavioral variety and document the training of a rough toothed porpoise (*Steno bredanensis*) in a task involving the performance of novel behaviors. The training program used by Pryor, Haag & O'Reilly consisted of two to four daily training sessions lasting five to twenty minutes each. Consecutive sessions were separated by a rest period of thirty minutes to one hour. Sessions began and ended with a context cue of a ringing bell to signal the beginning and end of each session. The authors note that the position of the trainers during sessions may also have acted as a context cue in addition to the bell.

During each session, the context bell would ring indicating to the subject that the session had begun and opportunity for reinforcement was present. In each session there was a trainer and there were two observers who verified each correct response. The trainer would give the cue, wait for the subject to offer a behavior, and assess if behavior met the criterion for reinforcement. This criterion was any behavior the subject exhibited that had not been previously offered in a training session (1969). When the subject offered a behavior matching the criterion, a whistle

bridge and reinforcement was given after which the trainer would wait for another behavior. Training would repeat in this manner until the trainer terminated the session with another ring of the bell, indicating the end of the session.

After thirty-two sessions the researchers concluded the training when the subject exhibited at least one novel behavior in the majority of seven consecutive sessions. Additionally, they reported that the novel behaviors became increasingly complex, making documentation of new behaviors increasingly difficult (Pryor et. al 1969).

The study by Pryor et al. (1969) reported that in total, sixteen novel behaviors were exhibited by the subject throughout training. Of these documented behaviors, five behaviors were known to be displayed as part of the natural repertoire of dolphins; four behaviors, while not exhibited in natural populations, have been demonstrated in other captive populations by shaping; three behaviors occur naturally in only one species of the subject's genus *Stenella* and had not been previously observed at the training facility; and four had never been observed in either captive nor wild species of the dolphin. The researchers suggest that though it is possible for all sixteen novel behaviors seen during the training to occur outside of training, the subject exhibited behavior to the cue outside of normal species behavior. Thus, a technique consisting of positive reinforcement for different, yet naturally occurring behaviors was sufficient to increase the likelihood of the subject to offer new behaviors. This seminal research influenced other researchers or trainers to pursue the notion of "the creative porpoise".

In an unpublished master's thesis, Braslau-Schneck (1994) utilized a population of two bottlenose dolphins (*Tursiops truncatus*) who had previous training with the concept of offering "different" behaviors when cued to do so i.e. offering each behavior exactly once during the present session, avoiding repeats, as well as previous training in performing behaviors in

tandem with each other, e.g. breaches from the water in a synchronous fashion. These two concepts of "different" and "tandem" behaviors were trained to the dolphins individually. The researcher reported the response of the two subjects when exposed to a cue for "tandem different". In report, the two subjects each offered the same "different" behavior in tandem, suggesting that not only did the subjects possess a previously established understanding of "different" as well as tandem behaviors, but could combine these concepts and coordinate with each other to produce different, tandem behaviors together.

More recently, Kuczaj and Eskelinen (2014) conducted a variation of Pryor and colleagues' study. There were a few key differences from Pryor's study. Instead of training novel behaviors, the researchers asked their dolphins to "vary" their behaviors during a series of training sessions. Three bottlenose dolphins were reinforced if they performed variations on the previous behavior or if they offered behaviors not previously seen during the current session after being given the "vary" cue.

The researchers observed that over the time of the training, subjects progressed from responding with simple behaviors (e.g., whistles and pectoral flipper waves) to offering increasingly complex behaviors containing multiple simultaneous body movement, vocalization or other such behavioral offerings referred to by the authors as components, such as waving pectoral fins whilst simultaneously whistling. The criteria for mastery of the task was defined as four consecutive training sessions in which each session contained at least three consecutive varied behaviors. Two subjects were reported to achieve this criteria after sixty-nine training sessions, while the third was reported to reach the milestone at seventy-four sessions. Once criteria for mastery of the task was met by each subject, a series of test trials for "vary" were

conducted with each subject in order to observe possible strategies used by the subjects to complete the task.

During these test trials, each subject was assessed separately in trials consisting of a trainer offering the "vary" signal and recording the subject's response, and if the response fit the criteria of being a different response to any previously offered during the present trial. On a correct response, this process of offering a cue and recording the subject's response was repeated. Trials continued until the trainer ended the session, or at which point the subject did not offer a correct, new response to the cue. These test trials were recorded on GoPro digital cameras, and each behavior was coded for analysis. This coding consisted of recording each behavior the subject responded to the cue with, as well as any and all components of an offered behavior.

It was reported that when given the 'vary' cue during test trials, the subjects were able to offer varied behaviors successfully in over ninety percent of their responses to the cue. This is a shift from their traditional training of the one to one ratio of cues to behavioral responses. By examining these test trials and the behaviors offered to the cue, the researchers concluded that their subjects understood the concept of "varying" their behavior when prompted, as well as the concept of not offering the same behavior twice during a session.

Kuczaj and Eskelinen (2014) suggested that the complex behaviors performed by dolphins were often combinations of other, simpler behaviors called components. For example, a dolphin would whistle and wave its pectoral fins simultaneously, a combination of two simple components resulting in a behavior that had not been previously offered. The number of components used to create each new behavior varied across subjects. For example, one subject

offered a maximum of ten components in one instance to create a behavior, while another offered six components in their most complex offered behavior.

Kuczaj and Eskelinen (2014) described a variety of possible strategies the dolphins employed, as it appeared that the different dolphins used different strategies. One strategy employed was labeled as "building", in which the dolphin would repeat the behavior previously performed and add a component to it. For example, if a dolphin offered a vocalization during the preceding trial, the current trial may consist of a vocalization while simultaneously offering a pec wave. Another strategy described by the researchers was labeled "deconstruction" in which a performed behavior might only exhibit a portion of the preceding behavior. For example, the present behavior may be a vocalization, whereas the previously offered behavior may have been a vocalization with a pec wave. The authors noted that the dolphins may utilize one, multiple, or sometimes no apparent strategy to successfully complete each trial, though no subject used one strategy exclusively.

Lastly, Kuczaj and Eskelinen (2014) examined the amount of energy the dolphins used to perform behaviors. The energy expenditure of each behavior was categorized as either low (e.g. a stationary pec wave), medium (e.g. swimming at a moderate pace), or high energy (e.g. a leap fully breaching the water). The behaviors exhibited by two of the three dolphins were predominately low energy behaviors; one dolphin primarily utilized high energy behaviors during test trials. Medium energy behaviors were offered occasionally by all subjects.

Kuczaj and Eskelinen (2014) claimed that the findings of their study, coupled with the findings of Mercado, Murray, Uyeyama, Pack and Herman (1998), "support the notion that dolphins represent their past actions and are able to use these representations to either repeat or modify something they have done" (Kuczaj & Eskelinen, 2014, p. 75). They suggest that

personality of dolphins or individual differences in how the dolphins interpreted the task may have influenced how each subject completed the task. Lastly, the researchers called for further research on the variability of behaviors that occur within and between dolphins during certain tasks, such as communication, imitation, or play.

In 2014, a social group of nine bottlenose dolphins at the National Aquarium in Baltimore, Maryland, began being trained on the concept of "different". This training task was initiated for two purposes: for its use as a mental enrichment for the dolphins as well as for its use for public education. The National Aquarium strives to educate its visitors on the cognitive abilities of the dolphin during their daily public training demonstrations. Karen Pryor, who conducted the first study on innovate training was consulted at the onset of the training. The task trained was closely aligned with that reported in the Kuczaj and Eskelinen (2014) study in that the dolphins were required to exhibit a behavior not been previously displayed during the present session. Specifically, the dolphins were reinforced if they responded with a behavior different from those they previously did in the same session. Over the course of training, the dolphins in this study were observed by the trainers to exhibit novel/innovative behavior on occasion.

The methods used by the training staff were as follows:

Description of training procedures and criteria used to train by training staff at the National Aquarium prior to the onset of this study.

Training sessions occurred on a highly variable schedule for the focal animals of the study, Foster and Beau. These two dolphins were chosen for this study because the consensus of the trainers at the National Aquarium was that Foster and Beau demonstrated mastery of the task most convincingly compared to the others in the social group. Foster had a total of 121 training sessions occurring over eighteen months; Beau had a total of 132 training sessions occurring

over eighteen months. Training occurred at least three times per week (Mondays to Fridays). Training took place at a variety of locations at the side of the exhibition pool as well as both habitat pools. Each session involved a trainer, a person who recorded the surface behavior of the responding dolphin with a video recorder, and a person who documented the responding dolphin's behavior on a datasheet. At the beginning of each trial, the "innovate" gestural signal, a hand clap directly above the head with arms fully extended, was given to the dolphin. The trainer waited for a behavioral response from the subject. On the initial trial of every session, any behavior given by the subject in response to the signal was reinforced. Each subsequent trial was judged on whether it was a behavior not previously offered during the present session. If the criterion was met, the subject was given a whistle bridge at the time when the behavior was exhibited followed by the delivery of food reinforcement when they returned to the trainer. If the criterion was not met, the trainer could either give the signal once again, or wait until another behavior was offered; if the criterion was met, the trainer would reinforce the correct response (see Figure 1).

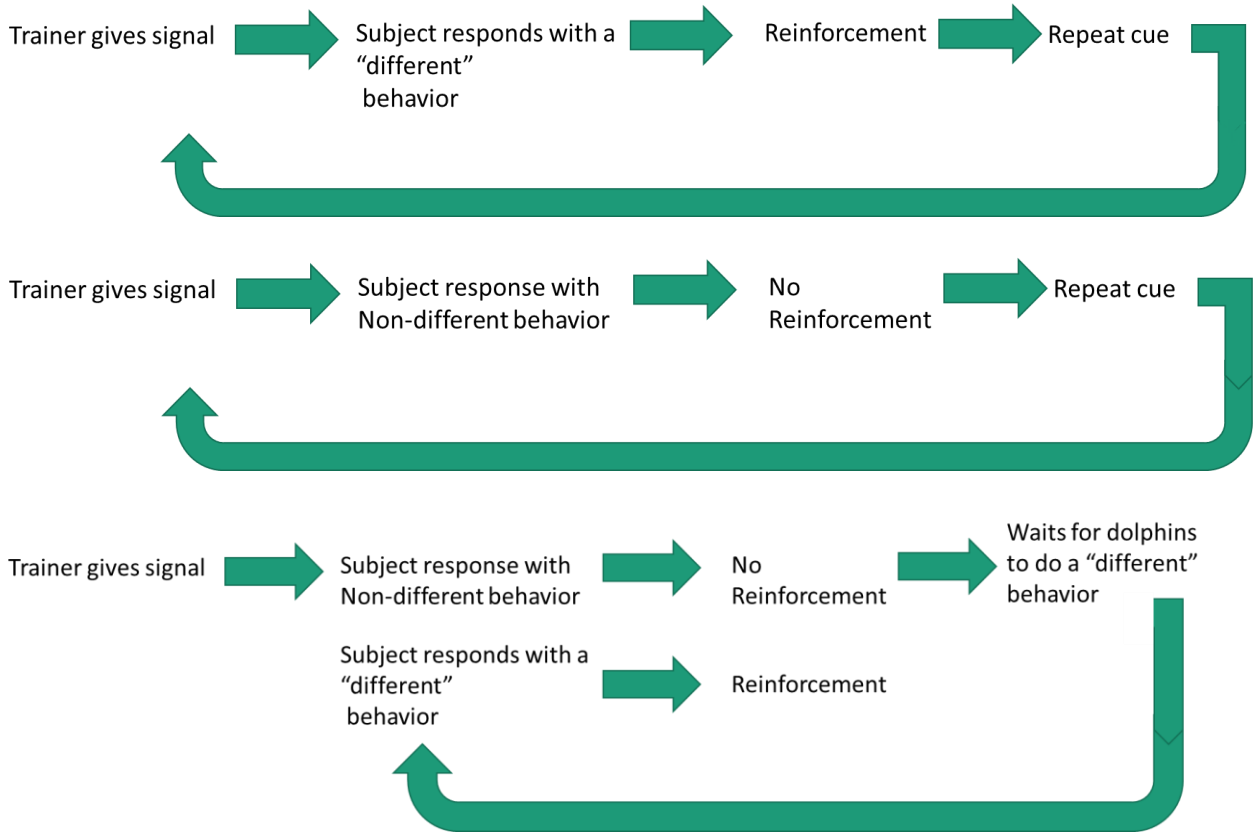


Figure 1. Structure of each "different/innovate" session

On some occasions, the trainers would reward some responses that, although "different", are frequently offered by the dolphins, with various toys or tactile stimulation. These are thought to be less desirable reinforcers and are utilized to break periods in which the dolphins routinely offered similar behaviors, a pattern documented by Pryor and colleagues, and to encourage further "different" behaviors (1969). Sessions were terminated either on reinforced behaviors or when aggression or disinterest from subjects was observed. The datasheet for each session contained information about each session including the trainer presiding over the current session, time of the session, time of day, date, subject, session number, when each gestural cue, or Discriminative Stimulus (S^D) was given (indicating a new trial), the behavioral response to the S^D, each subsequent behavior (if the first behavior after the S^D did not receive reinforcement),

total behaviors given per session, the type and amount of reinforcement given when applicable, any toys used with behaviors, and miscellaneous notes. After data collection during the training ended, demonstrations of "different" continued, as an educational program to demonstrate the cognitive abilities of the dolphins to visitors of the NA on a weekly basis.

The training afforded the opportunity for me to 1) analyze the data recorded during the first eighteen months to look for changes in performance over time, 2) examine training factors present that could shape the behavior and learning process of the task, 3) find possible strategies utilized by the dolphins to succeed in the task, and 4) collect new data approximately three years later, to compare with training in order to investigate if there was improvement, stability or decline in the subject's ability to succeed in the task. I sought to document how the dolphins responded to the cue "different/innovate" and if they showed behavior that suggested that they had learned the abstract concept of "different".

Methods

Subjects & Facilities

A total of eight bottlenose dolphins (*Tursiops truncatus*) were trained on the "different-innovate" cue: Nani (F) age 39; a calf of Nani's, Beau (M) age 9; Spirit (F) age 13; Chesapeake (F) age 22; a calf of Chesapeake's, Bayley (F) age 6; Maya (F) age 13; Jade (F) age 15; a calf of Jade's, Foster (M) age 7. The present study focused on the two male dolphins from this social group, Foster and Beau that were generally housed together in a separate but connected pool for husbandry reasons. Foster and Beau were selected at the discretion of the trainers, who suggested that amongst the group of dolphins, Foster and Beau had exhibited the highest aptitude for the task by the end of training. At the onset of the "different' training paradigm, Foster was 7 years old and Beau was 9 years old. Both animals were previously trained to respond to various

gestural cues for educational demonstrations, enrichment and husbandry at the National Aquarium. At the onset of training, Foster had been previously trained to perform 30 behaviors on specific cues, and Beau had been trained to do 22 behaviors on specific cues. All animals continued with normal care and feeding during the time of the training. No food deprivation was used during training or in this study.

All subjects were residents at the National Aquarium, Baltimore MD. The dolphins were housed in a 1.3M gallon pool, divided into four sections: one exhibition pool, two habitat pools and a medical pool, all interconnected. These facility dimensions were retrieved from a previous study conducted on the population at the National Aquarium (see Reid, Mann, Weiner & Hecker 1995).

Procedure

Phase I: Analysis of Phase I training sessions.

All datasheets for "different" training during the first year were collected from the training department at the National Aquarium for analysis. I analyzed the dolphins' responses to the cue "different" collected for Beau from October 29, 2014 to June 14, 2015 for a total of 132 sessions and the training data collected for Foster from October 25, 2014 to June14, 2015 for a total of 121 sessions. Datasheets were transcribed and entered into Microsoft Excel and IBM SPSS for analysis.

First Level Analysis: Frequency of Occurrence of Different Categories of Behavior. Each behavior was entered into a separate ethogram for each subject and was coded as one of four categories of behavior in the primary analysis: **I**, or "initial" denotes the first behavior offered during the present session. This behavior is always reinforced, and so is not considered in much of the analysis. **S**, or "same" denotes a behavior that has been previously exhibited during

the present session. **R**, or "repeat" is a behavior that is identical to the immediately preceding behavior. **D**, or "different" is the reinforceable behavioral category: it is a behavior not previously displayed during the present session. Frequencies of the categories of behavior S, R, and D were quantified across and between sessions.

Second Level Analysis: Categorization of show, previously documented non-show behaviors, or novel behaviors. If the dolphins were required to perform a different behavior than previously done in the session, they could accomplish the task by just performing a series of different show behaviors. To determine if this was a strategy employed, all behaviors were categorized as either 1) "show" 2) previously documented "non-show" or 3) novel behaviors. SHOW denotes show behaviors which referred to behaviors previously taught to the animals. Such behaviors were reinforced if they had not been previously displayed during the current session. Such behaviors were determined and categorized using trainer documents of previously trained behavior in each subject. N denoted "Novel" behaviors. These are behaviors not yet seen in the entirety of the initial training period, up to the current session. Such behaviors were identified by locating the first observed instance of every behavior recorded.

These categories of behavior were quantified across and between sessions. Since show behaviors may have fit criteria to be either different, same, or repeat behaviors during the preliminary analysis, frequencies of show behaviors as these other categories were observed and quantified. Novel behaviors are category "D" behavior in this analysis by definition. I also analyzed the initial behavior (1^{st} trial responses) that each dolphin exhibited in sessions to determine if they were show, non-show or novel behaviors.

Third Level Analysis: Examination of trials 2-6 of each session. To assess learning of the task, an analysis of every session was conducted in which the initial behavior was omitted,

and only the first behavioral response to the second, third, fourth, fifth and sixth S^D was observed. This was done to account for the possibility that memory might play a role in this task and as a function of this the task increases in difficulty with every behavior given regardless of their understanding of the task (see Kuczaj & Eskelinen 2014 for details; Mercado, Murray, Uyeyama, Pack, & Herman 1998). Using this paradigm, I quantified the frequencies of occurrence of the categories of behaviors in these trials.

Fourth Level Analysis: Quantification of "runs" of correct responses in trials # 2-6. To access if and when the dolphins developed the concept of "different", I quantified the frequency of occurrence and length of "runs" of consecutive correct responses to the cue. I used a binary scale of one and zero representing the presence or absence of a category "D" behavior in the present trial. A run in this context was operationally defined as three to five consecutive categories "D" trials all occurring within one session. Runs were obtained by adding a binary (1/0) category to every behavior, indicating if it was a category "D" behavior or either an "S" or "R" behavior. After excluding the first trial of each session, I concatenated all responses in each quarter and examined consecutive "different" behaviors and utilized a runs test of randomness using those lists (Wald & Wolfowitz 1943). I also assessed the consecutive "different" behaviors within sessions to verify if the present subjects reached a minimum of four consecutive sessions each consisting of one run of three or more "different" responses; the same criteria of mastery of the task as described by Kuczaj and Eskelinen (2014).

Phase II: Collection and analysis of data three years after initial training

The collection of data in Phase II was obtained as similarly as possible to Phase I's data collection. Collection began August 1, 2017 and ended August 31, 2017. In total, seven sessions

each were recorded for Foster and Beau, with some of the training sessions occurring during demonstrations for the public, while others occurred without an audience (see Table 1).

Table 1

Date of session and presence or absence of audience in each session during Phase II

Foster			**Beau**		
Session	**Date**	**Audience present/absent**	**Session**	**Date**	**Audience present/absent**
1	8.3	Present	1	8.2	Present
2	8.4	Present	2	8.16	Absent
3	8.8	Present	3	8.21	Present
4	8.25	Absent	4	8.23	Absent
5	8.28	Present	5	8.28	Absent
6	8.29	Absent	6	8.29	Absent
7	8.30	Present	7	8.31	Present

Sessions were structured in the same manner as the initial training, with the exception that the dolphin's behavior was not documented on a datasheet during the session. Instead, I videotaped each training session using a handheld Cannon Vixia HF R70. I stood adjacent to the trainer on the exhibition pool's center platform approximately two feet away from the trainer, on their righthand side and videotaped both the trainer and the dolphin to verify when the S^D was given by the trainer and to record the dolphin's responses to the cue. A second camera, a Cannon Vixia HF M500, was positioned in the audience area across the pool and also recorded the

dolphin's behavior in each session. Following each session, I reviewed the videos from both the camera at the training area and the camera overlooking the whole of the pool and used them to code the data onto datasheets having the same organization as those used for the initial training. All seven subjects participated in demonstrations of the task during the time Phase II data were collected, but only Foster and Beau's data were included in this study. Frequencies of each category of behavior were tabulated and compared to the results from Phase I.

Frequency of Occurrence of Different Categories of Behavior. Each behavior was entered into a new ethogram for each dolphin. These behaviors were categorized into the primary categories of behavior: I, S, R and D. These categories of behavior were quantified across and between sessions, and compared to the first level analysis of Phase I.

Examination of trials 2-6 of each session. To examine if an understanding of the task was maintained in Phase II, I conducted an analysis of the categories of behavior in the first response to the S^D during trials 2-6 of each session of Phase II. This analysis was compared with the same analysis of Phase I.

Quantification of "runs" of correct responses in trials 2-6. To assess if mastery of the task as defined by Kuczaj and Eskelinen (2014) was still present during Phase II, I once again examined runs of "different" responses to the S^D for each dolphin and compared this analysis to that of Phase I data.

Fourth Level Analysis: Lag sequential analysis. We conducted a lag sequential analysis (Bakeman and Gottman, 1997) on Phase II data as a means to determine whether and how "same' behaviors produced within each session were temporally related to each other.

Fifth Level Analysis: Observing construction strategies. We located and quantified observed instances of the construction strategy outlined by Kuczaj and Eskelinen (2014). To

identify instances of building, we examined each session from Phase II data and located via ethogram codes trials containing behavioral elements of the previous trial, such as a whistle in one trial (coded WH in ethogram) followed by a whistle with click in the subsequent trial (coded WHCL in ethogram) in the case of the "building" strategy.

Results

Phase I: Phase I training sessions.

First Level Analysis: Frequency of Occurrence of Different Categories of Behavior

Sessions consisted of ~6 to 9 trials (*Foster M = 8.98 SD = 5.79, Beau M = 6.35 SD = 3.96)*. The average duration of sessions was ~6 minutes (*Foster M = 396, SD = 183 seconds, Beau M = 361, SD = 128 seconds)*.

The preliminary analysis of the dolphins' responses to the new cue "different" revealed that Foster exhibited a total of 605 coded behaviors and Beau exhibited a total of 542 coded behaviors by the end of the first year of "different" training. Of these, 574 behaviors exhibited by Foster were not previously trained, and 509 behaviors exhibited by Beau were not previously trained. At the onset of training, Foster had previously been trained to perform 30 "show" behaviors on specific cues, and Beau had been trained to perform 22 "show" behaviors on specific cues as part of the educational program.

Throughout Phase I, Beau offered "different" behaviors in 34% of the trials, "same" behaviors in 32% of the trials, and "repeat" behaviors in 34% of the trials. Foster offered "different" behaviors in 45% of the trials, "same" behaviors in 43% of the trials, and "repeat" in 12% of the trials.

Second Level Analysis: Categorization of show, non-show behaviors, and novel behaviors. This categorization of the dolphins' responses to the cue "different" showed that both

dolphins exhibited show behaviors, non-show behaviors and novel behaviors during the Phase I training period. Show behaviors comprised 17% of Beau's responses and 31% of Foster's responses (Figure 2).

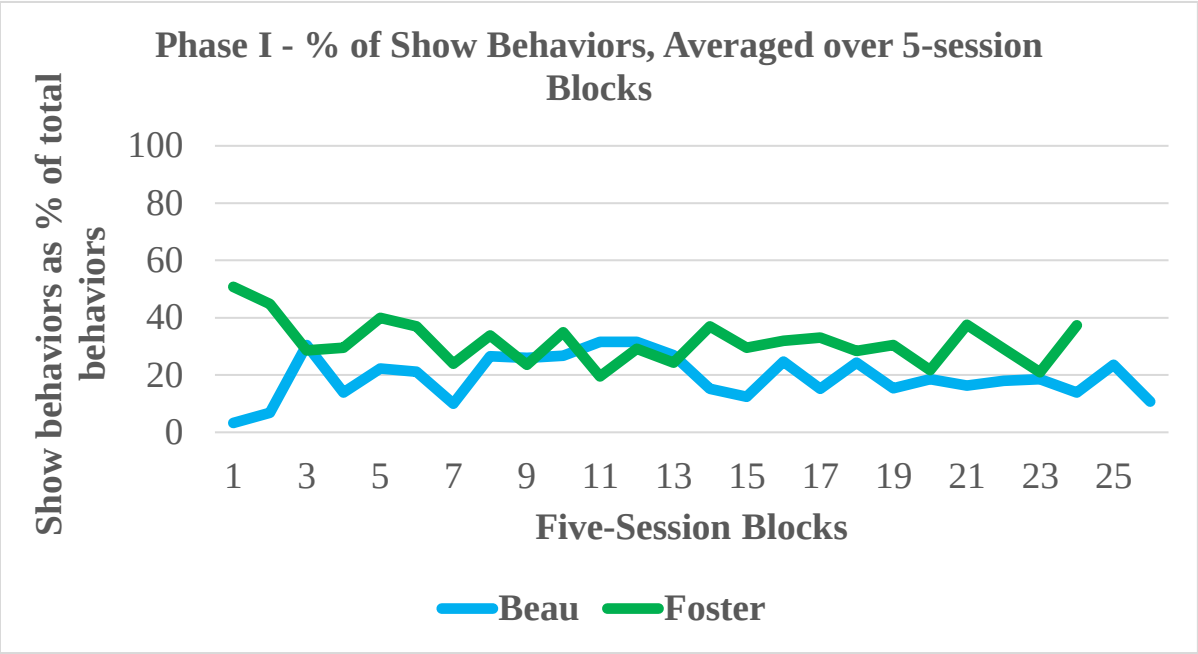


Figure 2. Percentage of novel behaviors exhibited in Phase I, averaged over 5-session blocks

Both dolphins exhibited novel behaviors, but the mean frequency of these behaviors did not exceed 20% of trials throughout Phase I, (*Foster M=12.20% SD = 7.59, Beau M=13.56% SD = 8.45)* (Figure 3). When compared to Pryor's criterion of exhibiting at least one novel behavior in the majority of seven consecutive sessions, both Foster and Beau achieved this within the first seven sessions of training. Further, Foster and Beau exhibited at least one novel behavior in all but eight of their respective sessions. Foster offered a maximum of sixteen novel behaviors in one session, while Beau offered a maximum of twelve novel behaviors in a session, on two occasions.

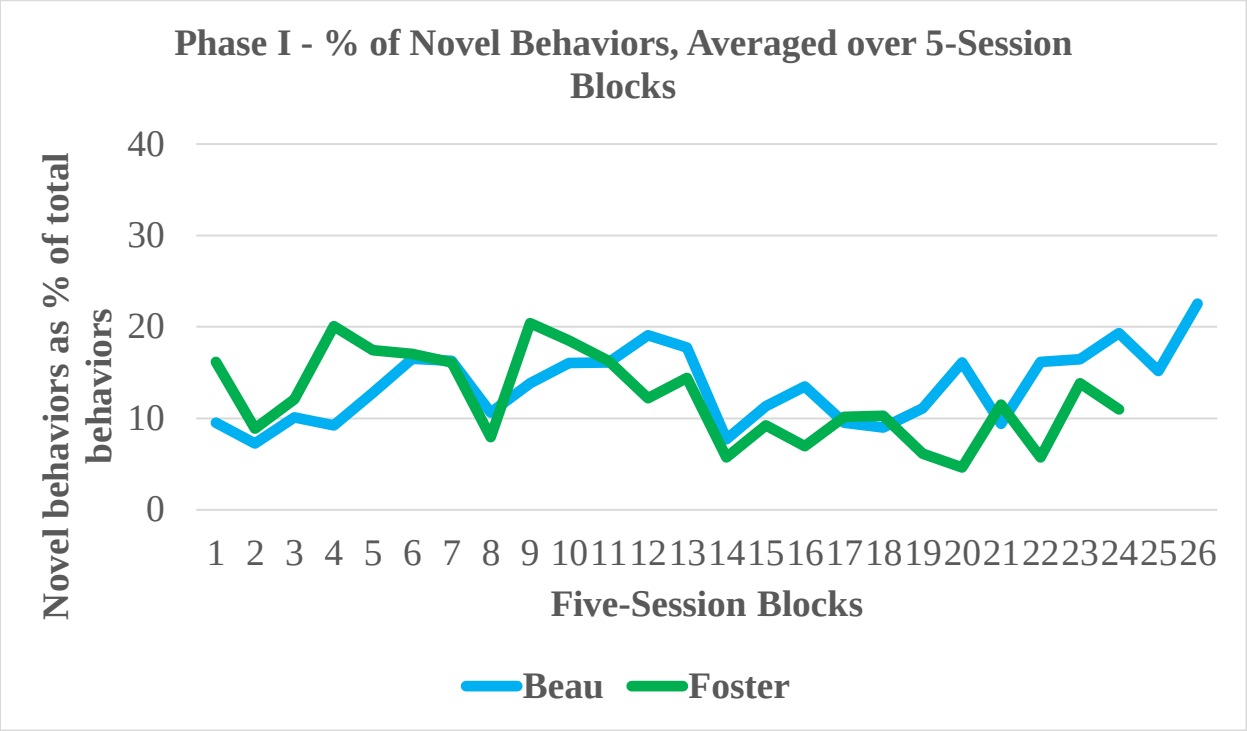


Figure 3. Percentage of novel behaviors exhibited in Phase I, averaged over 5-session blocks

The initial behaviors exhibited in sessions. Beau and Foster responded differently in their initial (first) responses to the "different" cue in sessions. During his 132 sessions in Phase I, Beau primarily exhibited "non-show" behaviors (show = 25%, novel = 15%, previously documented non-show = 60%). Out of his total 121 sessions in Phase I, Foster primarily exhibited the category of "show" behaviors as the initial behavior during each session (show = 82%, novel = 2%, previously documented non-show = 16%).

Third Level Analysis: Examination of trials 2-6 of each session. I analyzed each dolphin's first response to the S^D in trials 2-6 in each session in order to control for the possibility that the "different" task would increase in difficulty, regardless of the dolphins' understanding of the task, due to increasing memory requirements to recall past actions. The sessions in Phase I were divided into 1st, 2nd, 3rd and 4th quarters (~thirty consecutive sessions per quarter) in order to observe any frequency changes over time in any category of behavior. This

analysis indicated that category "S" was the dominant category of behaviors displayed by both subjects during the first quarter of Phase I.

Category "S" behaviors decreased and category "D" behaviors increased during sessions and category "D" behaviors were the dominant category in quarter 2 *(Foster D = 65.5%, Beau D = 49%).* Category "D" behaviors continued to be the predominant category of behavior exhibited during trials 2-6 in quarters 3 and 4 for both Foster *(quarter 3 = 66%, quarter 4 = 67.9%)* and Beau *(quarter 3 = 56.7%, quarter 4 = 67.9%).* Category "R" behaviors remained relatively stable across all quarters of Phase I for Foster (*quarter 1=21.5%, quarter 2=17.6%, quarter 3=17.3%, quarter 4=12.8%*) and for Beau (*quarter 1=15.2%, quarter 2=12.7%, quarter 3=24.4%, quarter 4=16.4%*) (Figure 4).

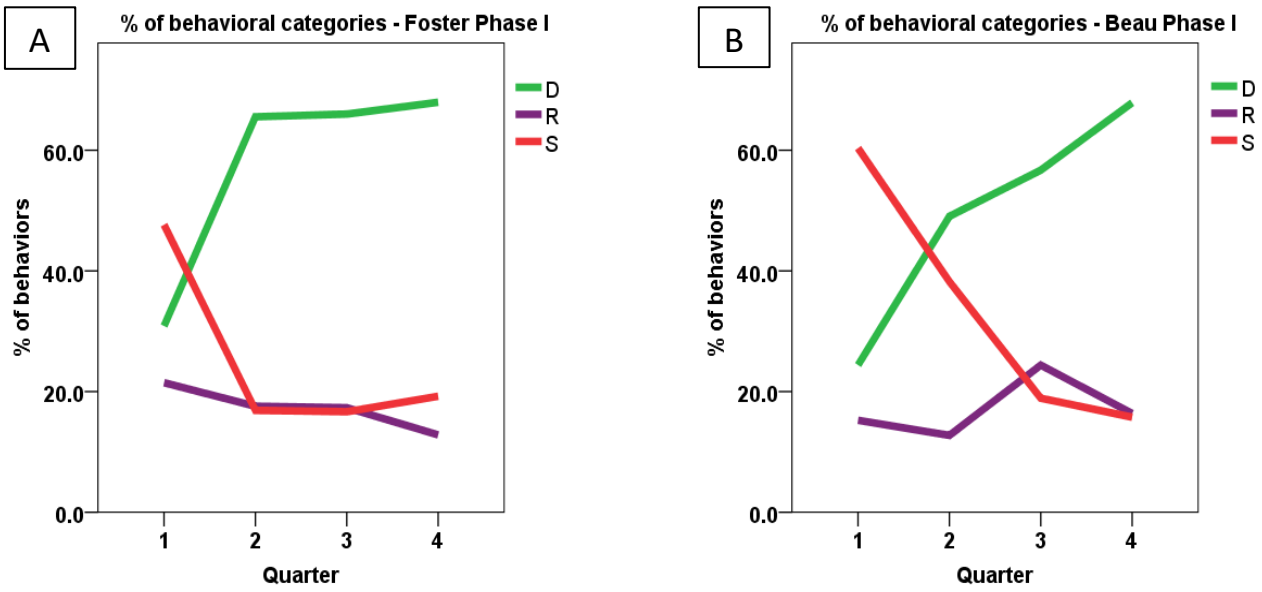


Figure 4. Percentages of category "D", "R", and "S" behaviors in the first response to S^D for trials 2-6 of each session, by quarter

Fourth Level Analysis: Quantification of "runs" of correct responses in trials 2-6. A two-tailed runs test for the randomness of a sequence of "different" or "not different" responses yielded significant results in three quarters of training for Foster (Q1 p=.309; Q2 p<.001; Q3 p<.001; Q4 p=.03) and in each quarter for Beau (Q1 p<.001; Q2 p<.043; Q3 p<.001; Q4 p<.001)

I conducted an analysis of the frequencies of occurrence of consecutive category "D" behaviors by each dolphin within sessions, termed "runs", to further elucidate whether the dolphins' were exhibiting an understanding of the task. Figure 5 shows the increase in the frequency of runs of novel and/or different behaviors by both dolphins over the course of Phase I. For both dolphins, the number of runs of different (correct) behaviors for trials 2-6 of 1st quarter sessions were low for both dolphins (Beau *n=3*), Foster (*n=5*) and progressively increased in both in the 2nd quarter sessions Beau (*n=10*), Foster (*n=13*), remained relatively constant (Beau *n=10*) or decreased slightly (Foster *n=12*) in 3rd quarter sessions, and increased in 4th quarter sessions (Beau *n=18*), Foster (*n=17*). While the numbers of runs per quarter increased progressively across Phase I, length of the runs indicated by the number of trials within runs did not increase for either subject (*Beau range: 3.3 - 3.9; Foster range: 3.7 - 4*).

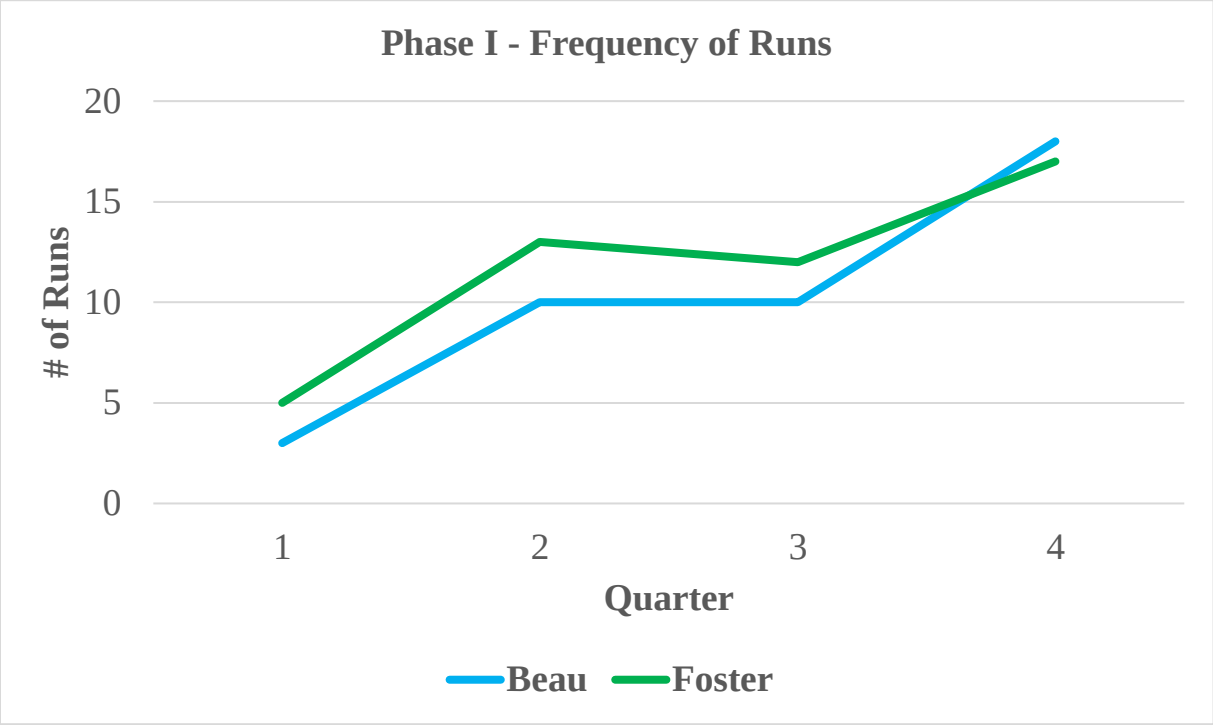


Figure 5. Frequency of Runs exhibited by Foster and Beau in each quarter of Phase I.

Phase II: Collection and analysis of data three years after initial training.

Sessions during Phase II lasted ~ three minutes (*Foster M = 155, SD = 32 seconds, Beau M = 201, SD = 102 seconds*) and contained ~ six trials per session (*Foster M = 6.36 SD = 3.80, Beau M = 5.94 SD = 3.59*).

First Level Analysis: Frequency of Occurrence of Different Categories of Behavior.

During Phase II, Beau responded with "D" behaviors in 61% of trials, "S" behaviors in 29% of trials, and "R" behaviors in 10% of trials, and Foster responded with "D" behaviors in 75% of trials, "S" behaviors in 16% of trials and "R" behaviors in 9% of trials (see fig. 6).

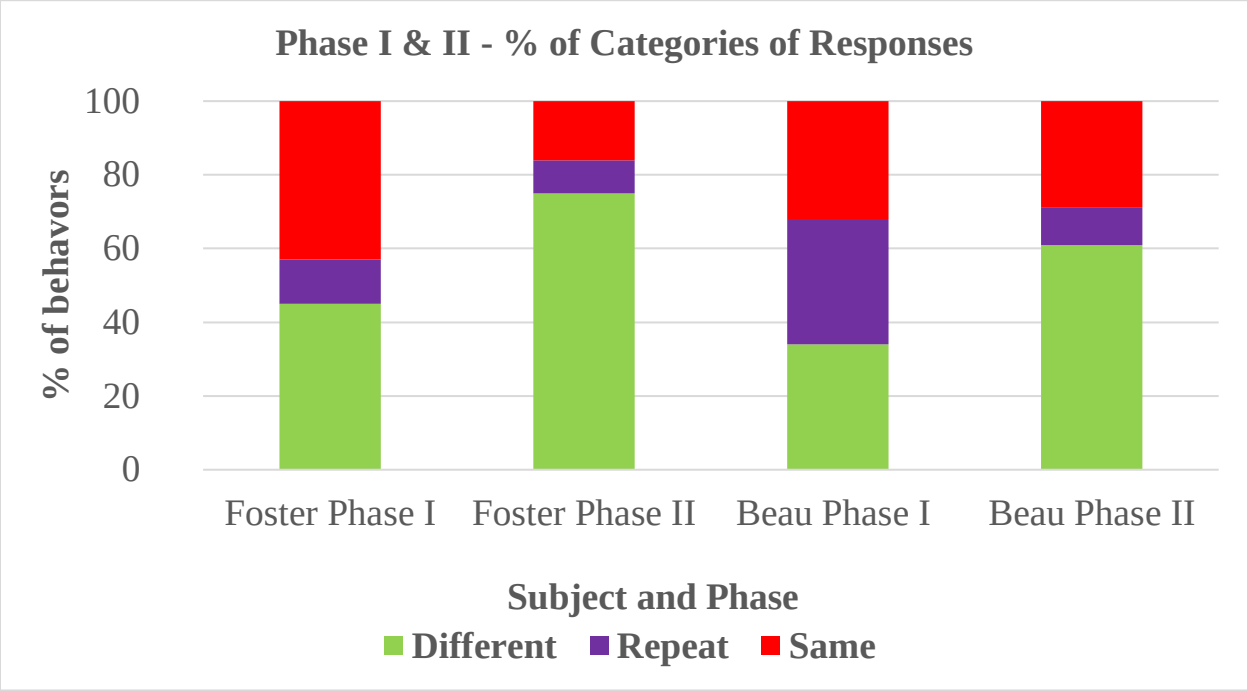


Figure 6. Percentage of categories of behavior compared between Phase I & II

Second Level Analysis: Examination of trials 2-6 of each session. Phase II analysis brought forth a similar pattern in the frequency of each category of behavior observed on the first response to the S^D in trials 2-6. Category "D" behaviors are still the most prevalent responses produced by both Beau and Foster as their first response to the S^D in trials 2-6 of each session (Figure 7.)

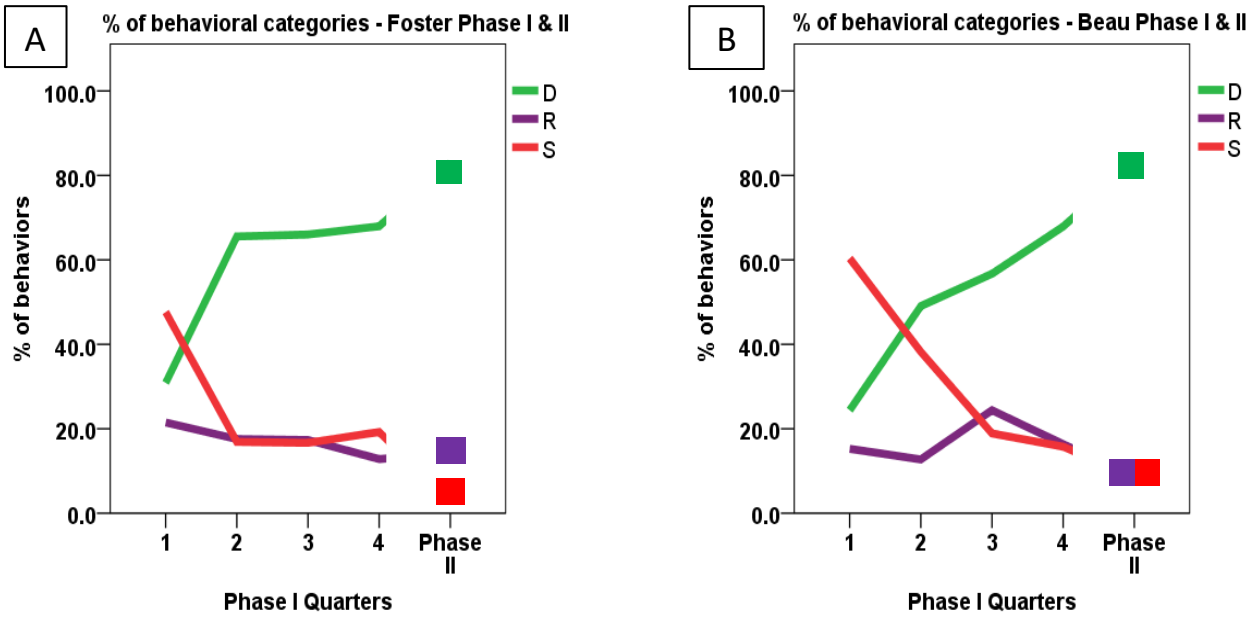


Figure 7. Frequencies of category "D", "R", and "S" behaviors in the first response to the S^D for trials 2-6 of each session in Phase II, compared to Phase I

Third level analysis: Quantification of "runs" of correct responses in trials 2-6. I quantified the frequencies of occurrence of consecutive category "D" behaviors "runs" by each dolphin within trials 2-6 of each session in the 7 sessions conducted in Phase II (Table 2). A runs test of randomness did not yield significant results for either dolphin, though using Kuczaj and Eskelinen's operational definition of a "run" (a minimum of four consecutive sessions each consisting of one run of three or more "different" responses), as the subsequent requirements for mastery of the "innovate" task and comparing to the analysis of runs for Phase I revealed that Beau has retained mastery during Phase II, while Foster does not.

Table 2

Number and Length of Runs by Beau and Foster in trials 2-6 in Phase II

	Beau			**Foster**	
Session #	**# of Runs**	**Run length**	**Session #**	**# of Runs**	**Run length**
1	1	5	1	1	5
2	1	5	2	1	3
3	1	4	3	1	5
4	1	4	4	0	0
5	1	3	5	1	4
6	0	0	6	1	4
7	1	5	7	1	4

Fourth level analysis: Lag sequential analysis. To examine patterns of behavior and possible strategies in use by the dolphins, a lag sequential analysis was conducted on the data from the 7 sessions in Phase II to determine when each dolphin produced the same or repeated a behavior previously performed within a session. The analysis revealed that category S (same) responses exhibited by Beau within sessions (n=35) primarily occurred 2 or 3 three trials after the first instance of that behavior and category R (repeat) behaviors were lower (n=6). In contrast, the category S responses exhibited by Foster within all sessions were low (n=32) and when he did a behavior more than once they were instances of category R behaviors occurring in the following trial (n=9).

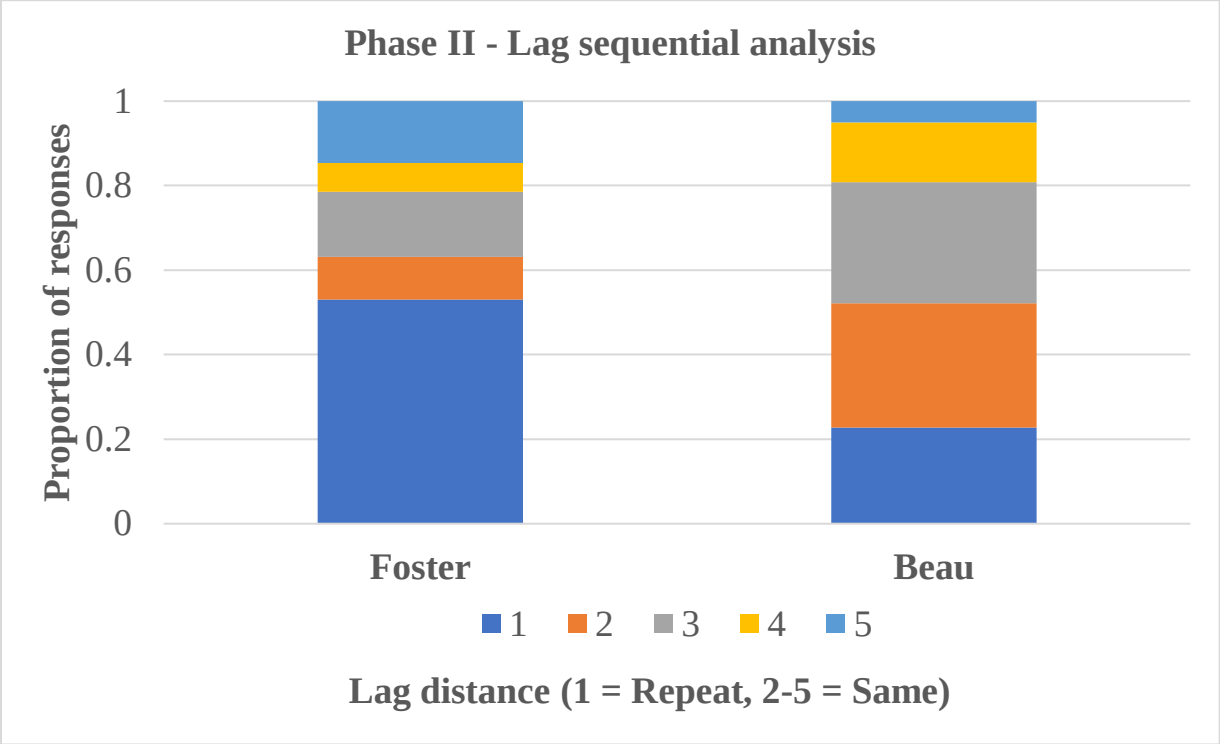


Figure 8. Lag sequential analysis Phase II

Fifth level: Observing construction strategies. While observing the specific behaviors the dolphins offered to complete the task during Phase II, the "building" of components to create new behaviors as outlined by Kuczaj and Eskelinen (2014) was occasionally observed. Subjects utilized components of behaviors previously offered such as buzzes and whistles with additional components not seen previously in the session. This strategy of building was observed more prominently in Foster's responses (n=6). Beau, while occasionally employing a building strategy during Phase II (n=3), appeared to employ other strategies, such as responding with behaviors previously performed in that session but doing them in different locations in the pool or with the addition of toys. In Phase II I did not observe instances of the deconstruction strategy.

Discussion

This study documented, analyzed and tracked the dolphins' behavioral responses and their trajectory of learning during training on the cue "do something different", and then during 7 sessions from demonstrations three years later. The analysis of the dolphins' behavioral responses during Phase I training indicated that the repertoires of behavior by both dolphins increased from their baseline trained 'show behaviors' to gestural cues (Foster, *n*=30, Beau, *n*=22) to those exhibited by the end of Phase I and Phase II the in response to the new cue "different" (Foster, *n*=574; Beau, *n*=509). Variety of behaviors increased, and an analysis of their behavior revealed changes in how they responded. This suggests not only further evidence of training for "different" behavior as an effective tool for increasing behavioral variety, but also suggests that the dolphins offered behavior outside of those which were shaped by trainers, and thus offered behaviors of their own design.

An analysis of autocorrelations proved significant, preventing statistical analyses such as chi squares which would assume independence of errors in the data. However, other tests that do not assume independence may prove interesting to pursue in future analyses.

During the first 30 sessions of Phase I, both Foster and Beau responded primarily with category S behaviors. This would be predicted, as all previously trained gestural cues in their repertoire correspond to a single behavior. In other words, if given the same gestural cue repeatedly, the dolphins were trained to do the same behavior repeatedly. Thus, given their training history and experience, it is likely and understandable that the dolphins repeated the behavior for which they had been previously reinforced. The shift from primarily offering R (repeat) or S (same) behaviors to D (different) behaviors when given the cue, appears to mark the beginning of their understanding of the new concept of "different" or "innovate"; this occurred

for both dolphins by the end of the first quarter of Phase I. However, it is important to note that these findings only represent the analysis of the data from trials 2-6 of each session.

When all trials in all sessions are viewed, as done in the analysis of a much smaller data set of Phase II, category "S" behaviors contribute to a least one third of all behaviors for both dolphins, similar to the findings reported by Pryor et al. (1969). The dolphin in the Pryor et al. study was described as becoming fixated on a selection of behaviors that they would utilize repeatedly in a trial. Pryor and colleagues suggest this may be a result of frustration at the inherent difficulty of the task, regardless of understanding (Pryor at el. 1969; per. comm).

The lag sequential analysis of category S and R behaviors for Phase II data indicated that Beau and Foster exhibited different patterns when they exhibited these categories of behavior. The analysis closely examined Same and Repeat behaviors, and located the previous occurrence during the session of the present, non-different response to the S^D. Using this analysis revealed that previously seen behaviors exhibited by Foster occurred on the immediately preceding trial, indicating more prominent "repeat" category behavior, than "same" during incorrect responses to the S^D. Non-different trials (same category behaviors) exhibited by Beau appear to most often appear two and three trials previously in the session. Foster and Beau's behavioral patterns appear similar to those reported by Pryor et. al. (1969), though this finding may lead to more questions regarding the role of memory in this task.

The analysis of runs adds further evidence that the dolphins developed the concept "different". Firstly, the statistical runs test for randomness proved significant in most quarters of Phase I, with the exception being the first quarter of Phase I for Foster. The significance was based on negative Z-values which indicates multiple instances of consecutive "different" behaviors in response to the S^D, more than which would be expected by chance. Further evidence

that the dolphins had a concept of different is that they showed an increase in the performance of "different" behaviors and a decrease in the performance of "same" behaviors between quarter one and two of training (see figure 4). The insignificant results of a statistical analysis of runs for Phase II data may be explained by an insufficient number of sessions, and so future testing may yield significance.

Second, I used the same operational definition of a "run" and criteria for mastery of the task "different" that was used in reporting a conceptual understanding of "vary" reported by Kuczaj and Eskelinen (2014). Here, a "run" required an animal to perform a minimum of three different behaviors without repeating a behavior. Four consecutive sessions, each containing at least one "run" indicated mastery of the task. In Phase I, Foster achieved this criterion in sessions 116 through 120 and Beau achieved the criterion in sessions 102 through 106, and again in sessions 127 to 131. In Phase II Beau continued to show mastery of the task in Phase II whereas Foster did not but this may be due to an insufficient number of sessions observed in Phase II to reliably show mastery. However, these runs analyses were conducted by observing trials 2-6 of each session. This analysis was limited to trials 2-6 to remove the initial trial from the analysis as well as to control for the task's increase in difficulty regardless of an understanding of the task due to possible memory constraints postulated by Kuczaj and Eskelinen (2014). By using these five trials, the dolphins are only able to offer a run of three to six consecutive "different" behaviors once per session. Therefore, it is possible that throughout Phase I, additional runs of three or more consecutive "different" behaviors are present in later trials during sessions containing more trials.

Instances of novel behaviors were present throughout Phase I, with both dolphins meeting Pryor's original success criteria of offering at least one novel behavior in the majority of

seven consecutive sessions immediately. This steady exhibition of novel behaviors adds support to Pryor et. al. (1964), that dolphins are capable of offering behaviors not created by traditional training methods such as shaping. It is important to note that while the present training paradigm did not train explicitly for the creation of "novel" behaviors, such offerings nevertheless emerged as a result of the training.

Instances of "building" were observed during Phase II, supporting the findings in Kuczaj and Eskelinen's report (2014). However, instances of the "deconstruction" strategy were not found during Phase II. The training practice of waiting for additional behavioral responses on failed trials may encourage the strategy of building, as the dolphins are given the opportunity to build on the previous trial in order to meet the criteria for reinforcement. This may explain the presence of the building strategy.

The findings of this study provide additional evidence in support of the findings reported by Pryor and et. al. (1964) on training for creative behavior. The findings also lend supporting evidence to the use of strategies documented by Kuczaj and Eskelinen (2014). These studies provide evidence for the capacity for behavioral innovation both in captive and wild populations.

Differences in speed of task acquisition may be due to inter-study differences. The use of multiple trainers in the present study may have contributed to the time it took to achieve mastery of the task by both dolphins. The use of a single trainer for the duration of training as opposed to the use of multiple trainers could have effects on the consistency and success of the training. The concept of "different" is a broad, abstract term and dependent on the immediate perception and judgement by different trainers working with the same dolphin. As a result, criteria for reinforcement may change slightly every session depending on which trainer is involved. A

single trainer presiding over every session can have one uniform definition of "different", and thus the criteria for reinforcement will remain the same throughout training.

During training of the task, it is imperative to keep the criteria for reinforcement as consistent as possible. For example, during the beginning of Beau's training in Phase I, spyhopping was a prominent behavioral response. In this case, the behavior may have been a general orienting behavior that was reinforced in sessions. Pryor, Haag and O'Reilly (1969) reported a similar pattern in which specific behaviors were repeated; the authors selected and shaped other behaviors in order to interrupt the persistence of repeated behaviors by their subject. While the trainers of the present study implemented this process as well, certain repeated behaviors were not reinforced on all subsequent sessions, regardless of whether they met criteria for reinforcement. This "rule change" occurred around session 20 of training. The discontinuance of reinforcement of a specific behavior such as a spyhop, that was reinforced in earlier sessions, could possibly confuse the dolphins and hinder acquisition of the task as well as result in the cultivation of frustrated behavioral responses. Therefore, in future training paradigms and research all behaviors other than overt aggression or sexual displays should be reinforced at all times, regardless of how often they are displayed, provided they meet criteria for reinforcement.

Since the term "different" is broad, abstract and often not consistently defined for human trainers, and it is likely that the semantics of the cue is abstract during training for the dolphins, the trainer's definition of "different" and the subject's "definition" of different may not be aligned, especially when considering multiple trainer worked with the dolphins. Thus, this task not only becomes a memory task, as suggested by Kuczaj and Eskelinen (2014), but also a communication task involving clarification of the requisite criteria, from the perspective of a transactional model of communication (Barnlund 1970). For example, during the training

process, when the trainer gives the cue “do something different”, the subject offers a behavior that it may perceives as “different”. The trainer then decides whether or not they agree that the behavior meets the criteria for reinforcement. If the two agree, reinforcement is administered, and the session continues. When the trainer decides not to provide reinforcement, it is possible that the dolphin thinks the behavior offered meets the criterion, while the trainer does not.

The “building” and “deconstruction” strategies outlined by Kuczaj and Eskelinen (2014) may account for some of these differences in perception of the meaning of the cue “different’ between trainer and dolphin. A trainer may not agree that a behavior is “different” if it contains a component previously seen within the session. Additionally, previously offered behaviors may be “different” from the dolphin’s perspective if performed for example, in a different location or in conjunction with a toy. These modifications to previously offered behaviors may be strategies used by the dolphins to complete the task. Thus, from a transactional communication perspective, the training context can be viewed as a dynamic set of transactions between the trainer and dolphin in which the semantics (meaning) of the cue may not be the same for the two interactants and thus the training process involves the synchronization of the behavior of both during which time the concept “different” is negotiated.

Dolphin species are prime candidates for the study of the genesis of new behaviors, according to Reader (2003), who proposed characteristics of species likely to “innovate”. He suggested innovation would mostly likely be found in species high in explorative natural behaviors, high numbers of individuals within the population that express neophilic tendencies, and the species’ ability to learn behaviors both socially and asocially.

Beyond studies that specifically have focused on training and testing the concept of innovate, evidence for an aptitude of creative behavior has been reported in other studies of

animal behavior and cognition (for full review, see Patterson & Mann, 2015; Kuczaj 2017). The propensity of social and asocial learning and information transmission between cetaceans is extensively documented (Whiten, 2007; Reiss and McCowan, 1993; McCowan et al, 2000; Lopes, Borger-Turner, Eskelinen, & Kuczaj, 2016). In neurological research, cetaceans possess brain-body ratios beyond that of chimpanzees, and second only to humans (Marino, 1998). Tool use has been observed and documented in the dolphin population of Shark Bay, Australia, where the population has utilized sponges as protective instruments in foraging through abrasive coral reefs (Smolker, Richards, Connor, Mann, & Berggren, 1997), perhaps indicating a propensity for neophilic and explorative behavior. Per the species identified by Reader (2003), good candidates with which to study innovative behaviors are primates, magpies, and dolphins. However, innovation research is not exclusive to these taxa. Studies on novel behaviors in fact begin with a series of observations on foraging tactics with finches opening milk bottles (Hinde & Fisher, 1951). Similar studies expanded to include taxa both within and outside of Reader's proposed groups including walruses, orangutans and guppies (Pryor, Wylie & Chase 2014; Laland & Reader 1999; Schusterman & Reichmuth, 2008; Russon, Purwo, Ferisa & Handayani 2010).

The present research provides further support for the capacity of dolphins to understand the concept of 'different'. Furthermore, our results provide additional evidence that dolphins possess the ability to mentally represent their own past actions and use this metaknowledge to enable them to either repeat past behaviors when asked (Mercado, Murray, Uyeyama, Pack & Herman (1998), or avoid repeating past behaviors and/or use components of past behaviors to create novel behaviors (Kuczaj and Eskelinen 2014). Our findings provide insights and suggestions for future studies on innovate training paradigms. I suggest that fewer trainers and clearly defined behavioral criteria may be optimal for learning this task. From a communication

perspective, future studies might view the interactive training paradigm as a dynamic system of communicative transactions between two intelligent species in which their respective definitions of the abstract concept “different” are negotiated.

References

Andrew Witen, C. P. v. S. (2007). The evolution of animal ‘cultures’ and social intelligence. Philosophical transactions of the royal society, 362, 17. doi:10.1098/rstb.2006.1998

Bakeman, R. and Gottman, J. (1997). Observing Interaction: An Introduction to Sequential Analysis Cambridge University Press, Cambridge.

Braslau-Schneck, S. (1994). *Innovative behaviors and synchronization in bottlenose dolphins* (Doctoral dissertation, University of Hawaii at Manoa).

Hinde, R. A., & Fisher, J. (1951). Further observations on the opening of milk bottles by birds. British Birds, 44(12), 393-396.

Kuczaj, S. A. (2017). Animal creativity and innovation. *APA Handbook of Comparative Psychology. Washington, DC: APA. http://doi. org/129*.

Kuczaj, S. A. II., Eskelinen HC. 2014 The ‘creative dolphin revisited: what do dolphins do when asked to vary their behavior. *Anim. Behav. Cogn, 1,* 66-76.

Laland, K. N., & Reader, S. M. (1999). Foraging innovation in the guppy. Animal Behaviour, 57(2), 331-340.

Lopes, M., Borger-Turner, J., Eskelinen, H., & Kuczaj, S. II.(2016). The influence of age, sex, and social affiliation on the responses of bottlenose dolphins (Tursiops truncatus) to a novel stimulus over time. Animal behavior and cognition, 3(1), 32-45.

McCowan, B., Marino, L., Vance, E., Walke, L., & Reiss, D. (2000). Bubble ring play of bottlenose dolphins (Tursiops truncatus): implications for cognition. Journ of Comp Psychol, 114(1), 98.

Mercado, E., Murray, S. O., Uyeyama, R. K., Pack, A. A., & Herman, L. M. (1998). Memory for recent actions in the bottlenosed dolphin (Tursiops truncatus): Repetition of arbitrary behaviors using an abstract rule. *Animal Learning & Behavior*, *26*(2), 210-218.

Patterson, E. M., & Mann, J. (2015). Cetacean innovation. In *Animal creativity and innovation* (pp. 73-125).

Pryor, K. W., & Chase, S. (2014). Training for variability and innovative behavior. *International Journal of Comparative Psychology*, *27*(2).

Pryor, K. W., Haag, R., & O'Reilly, J. (1969). The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of behavior*, *12*(4), 653-661.

Reader, S. M. (2003). Innovation and social learning: individual variation and brain evolution. *Animal biology*, *53*(2), 147-158.

Russon, A. E., Kuncoro, P., Ferisa, A., & Handayani, D. P. (2010). How orangutans (Pongo pygmaeus) innovate for water. *Journal of Comparative Psychology*, *124*(1), 14.

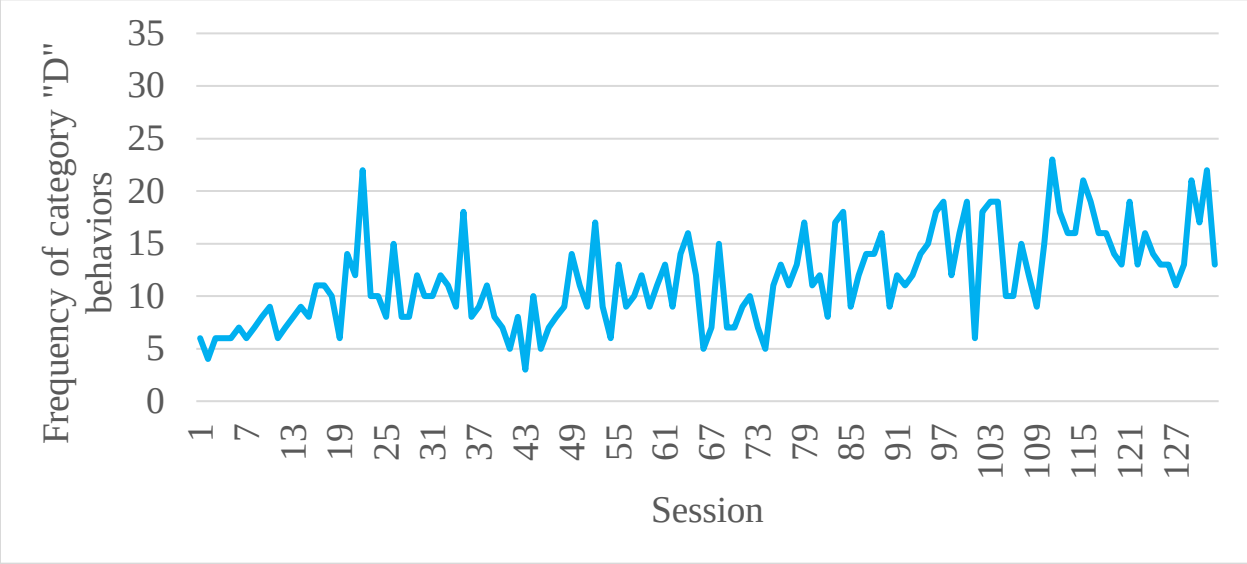
Schusterman, R. J., & Reichmuth, C. (2008). Novel sound production through contingency learning in the Pacific walrus (Odobenus rosmarus divergens). *Animal Cognition*, *11*(2), 319-327.

Skinner, B. F. (1951). How to teach animals. *Scientific American*, *185*(6), 26-29.

Smolker, R., Richards, A., Connor, R., Mann, J., & Berggren, P. (1997). Sponge carrying by dolphins (Delphinidae, Tursiops sp.): a foraging specialization involving tool use? Ethology, 103(6), 454-465.

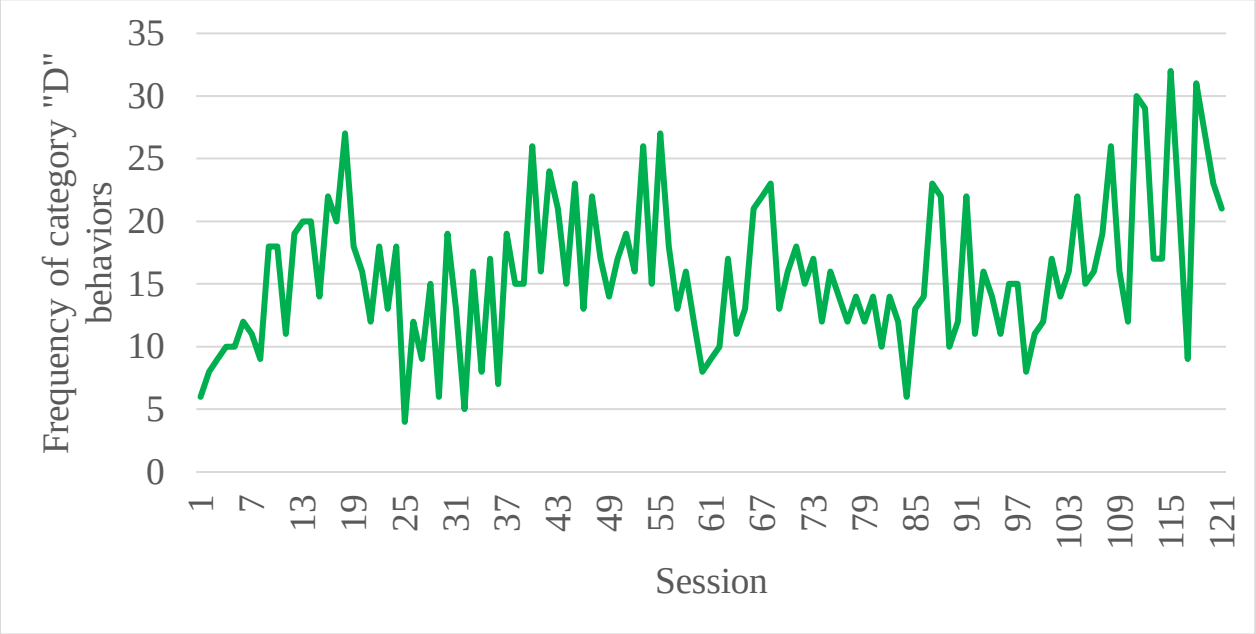
Appendix A

Frequency of category “D” behaviors in all responses between sessions. Beau Phase I.



Appendix B

Frequency of category “D” behaviors in all responses between sessions. Foster Phase I



Appendix C

Components and modifiers utilized to create novel behaviors, Phase I

Component list	Definition	Body Part Modifier	Definition	Location Modifier	Definition
360	360	B	Back	CD	center deck
APP	applause	BLY	belly	CSO	Central slide out
ARCH	arch	D	dorsal	D	deck
BB	bubble	DME	(dorsal) mermaid	FSO	far slide out
BOW	bow	FL	fluke	GLS	glass
BRCH	breach	F	face	GT	gate
BRLRL	barrel roll	H	Head	LG	ledge
BRTH	breath	LPC	left pec	L	to left
BUP	back up	ME	melon	LSO	left slide out
CBDY	"C" body position	OBH	on blowhole	MP	medical pool
CIRSW	circle swim	PC	Pecs	OD	On deck
CL	clicks	RO	rostrum	OL	on ledge
CM	closed mouth	RPC	right pec	OP	outflow pipe
D	(body part) down	SD	side	PIT	pit
DBPS	double back pec slap	TL	tail	R	to right
DBPW	double back pec wave	V	ventral	SSO	Show slide out
DRIB	dribble			T	travelling
DT	deck target			USO	under slide out
DU	dunk			UW	underwater
DV	dive			WI	window
DV	dive			WS	wetspot
DW	down				
FLNG	fling				
FLOAT	float				
FLP	flip				
FTW	forward tail walk				
GL	glide				
GRB	grab				
GRBTY	grab toy				
HAPP	high applause				
HLO	haul out				

HLOA	haul out attempt
HORSP	horizontal spin
HS	headstand
HSHK	Headshake
HULA	hula
HYPL	hydroplane
L	look
LAFOSTER	look at foster
LAY	layout
LAYOBTM	lay on bottom
LOB	fluke lob
LOB	lob
LP	loop
LSW	lateral swim
MB	body part movement (back)
MER	mermaid
MO	(body part) motion
MUPD	(body part) movement up and Down
O	On
OM	open mouth
PCO	pec on (location)
PK	peek
PLAY	play
POP	Pop up
POR	Porpoise
POSE	pose
PRS	(body part) present
PULLTYIP	pull toy into pool
PUSHTY	push toy
RAM	ram
RAS	raspberry
RL	roll
RUB	rub
S	(body part) stand
SH	Spyhop
SHARK	shark
SHK	(body part) shake
SIGWH	signature whistle
SLO	slide off
SLP	slap
SLSW	slow swim
SO	stand (on body part)
SOM	somersault

SPIT	spit
SPL	splash
SPN	spin
SPSW	spin swim
SQ	squawk
ST	Station
STRUT	strut
STW	stationary tail walk
SUCTY	suction toy
SW	swim
SWSTR	swim "strange"
T	target
TCHTY	touch toy
THR	(body part) thrash
TKO	take off
TOSSTY	toss toy
TR	travel
TRN	turn
TRNUPD	turn upside-down
TTCH	tummy touch
TWRL	twirl
TY	toy
UP	fluke up
UP	(body part) up
W	With
WV	(body part) wave
WH	whistle