Small and large number discrimination in goldfish (*Carassius auratus*) with extensive training

Caroline M. DeLong [a],[*], Stephanie Barbato [a], Taylor O'Leary [a], K. Tyler Wilcox [b]

[a] Department of Psychology, College of Liberal Arts, Rochester Institute of Technology, 18 Lomb Memorial Dr., Rochester, NY 14623, USA
[b] Department of Applied Statistics, College of Science, Rochester Institute of Technology, 85 Lomb Memorial Dr., Rochester, NY 14623, USA

ARTICLE INFO

ABSTRACT

Previous studies on relative quantity discrimination in birds and mammals with training procedures have employed hundreds or thousands of trials whereas studies with fish typically use dozens of trials. The goal of this study was to examine whether more extensive training improves the performance of fish tested on stimuli in the small (<4) and large (>4) number range. Goldfish were trained with dot array stimuli using the ratio 0.5 (2 vs. 4, 6 vs. 12) across two blocks of training sessions with a total of approximately 1200 trials. They were tested after each block of training sessions with the ratios 0.33 (1 vs. 3, 5 vs. 15), 0.5 (2 vs. 4, 6 vs. 12), and 0.67 (2 vs. 3, 10 vs. 15). Performance exceeded 90% correct on both test blocks. Accuracy was not affected by manipulating the surface area, density, or space of stimuli. Performance was best on the ratio 0.5 in test block 1, but ratio-independent in test block 2. There was no difference in performance in the small vs. large number range across the study. These results suggest that fish given extensive training can achieve accuracy on a numerical task comparable to well-trained birds, humans, or non-human primates.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The ability to assess numerical information and discriminate between two quantities is advantageous in many contexts. It allows animals to optimize foraging by comparing the number of prey items in different patches (Lucon-Xiccato et al., 2015), engage in effective predator defense by joining a larger social group (Hager and Helfman, 1991), and make informed decisions about whether to engage in fights based on assessing the number of opponents (McComb et al., 1994). Numerical abilities are present in a wide variety of animal species that include mammals, birds, amphibians, fish, and insects (see Agrillo and Bisazza, 2014; Reznikova and Ryabko, 2011 for reviews). For example, numerical competence has been found in monkeys (Beran and Parrish, 2013; Cantlon and Brannon, 2007), baboons (Cantlon et al., 2015), chimpanzees (Boysen and Berntson, 1989; Tomonaga, 2008), orangutans (Vonk, 2014), dolphins (Jaakkola et al., 2005; Kilian et al., 2003, 2005; Yaman et al., 2012), elephants (Irie-Sugimoto et al., 2009; Irie and Hasegawa, 2012), raccoons (Davis, 1984), cats (Pisa and Agrillo, 2009), pigeons (Emmerton and Remmer, 2009), chicks (Rugani et al., 2008), parrots (Pepperberg, 1994), mosquitofish (Agrillo et al., 2007, 2008), guppies (Piffer et al., 2012, 2013), salamanders (Uller et al., 2003), honeybees (Srinivasan, 2010), spiders (Rodriguez et al., 2015), and ants (Reznikova and Ryabko, 2011).

Humans and many non-human animals share the ability to compare quantities of items in two sets and determine which set has more items (e.g., compare 3 dots to 6 dots). This ability is not dependent on language or culture and is present in human infants (Izard et al., 2009; Xu and Spelke, 2000) as well as human adults tested under conditions that prevent counting (Agrillo et al., 2015; Cordes et al., 2001). Some researchers suggest that two different systems for non-symbolic numerical representation exist. The approximate number system (ANS) is used for representing large numbers (>4), and the object tracking system (OTS) is used for the precise parallel representation of 1–4 items (Trick and Pylyshyn, 1994; Xu and Spelke, 2000). The process of rapidly and accurately assessing the number of items in a small set (<4) is called subitizing (Mandler and Shebo, 1982; Piazza et al., 2002).

In a relative quantity judgment task, individuals are typically presented with a number of comparisons representing different ratios (e.g., 6 vs. 12 items = 0.50 ratio, 15 vs. 20 items = 0.75 ratio). If performance is ratio-dependent (i.e., shows a ratio effect), then accuracy will decrease as the numerical ratio between the smaller and larger set increases in accordance with Weber's Law (e.g., better performance on 0.50 ratio than on the 0.75 ratio). In contrast, no difference in accuracy between numerical ratios indicates ratio-independent performance (i.e., no ratio effect). The presence of ratio sensitivity is considered to be a hallmark of the ANS, whereas ratio

insensitivity indicates the recruitment of the OTS in human adults (Feigenson et al., 2004).

To date, numerous studies of relative quantity judgements in non-human animals reveal several different patterns of performance regarding ratio dependence in the small vs. large number range which may indicate the presence of either one or two numerical representation systems. Some studies suggest the presence of both the ANS (large numbers >4) and OTS (small number range 1–4), whereas others suggest that the ANS is activated over the entire numerical range. Research with rhesus monkeys (Cantlon and Brannon, 2006, 2007), olive baboons (Cantlon et al., 2015), chimpanzees (Tomonaga, 2008), bottlenose dolphins (Jaakkola et al., 2005), and three-spined stickleback fish (Mehlis et al., 2015) has found the ratio effect occurs in both the small and large number range. Two recent studies on Asian elephants report a lack of ratio dependence in both the small number range (Irie-Sugimoto et al., 2009) and the large number range (Irie and Hasegawa, 2012). However, Perdue et al. (2012) found a ratio effect across both ranges for African elephants. A single analogue magnitude system is implicated when the data shows ratio sensitivity over both number ranges or ratio insensitivity over both number ranges.

In contrast, there is evidence to suggest that a distinct OTS system operates in the small number range when ratio insensitivity is seen with small numbers and ratio dependence is found with large numbers (e.g., *canines*: Bonanni et al., 2011; *dolphins*: Yaman et al., 2012; *robins*: Hunt et al., 2008). A group of recent studies with fish provide data consistent with two number systems (*guppies*: Agrillo et al., 2012c, 2014; Bisazza et al., 2010; Piffer et al., 2012; *mosquitofish*: Agrillo et al., 2007, 2008; *angelfish*: Gomez-Laplaza and Gerlai, 2011a,b). For example, Agrillo et al. (2012c) showed that both undergraduate students viewing dot arrays and guppies in a shoal-preference task displayed ratio-dependent performance with large numbers but not small numbers. Piffer et al. (2012) found that guppies could discriminate between quantities in which stimuli were within the large number range (5 vs. 10) or within the small number range (3 vs. 4), but failed at quantity comparisons across the small-large boundary (3 vs. 5).

It is important to keep in mind that studies of relative quantity judgment vary widely in methodological approach and procedure, and these differences can impact the results. Spontaneous choice tasks and training procedures are the two main approaches; stimuli (visual shapes, auditory tones, food, shoaling companions), stimulus presentation (simultaneous, successive), presence or absence of distractors, number of training or testing trials, and presence or absence of feedback can vary (Agrillo and Bisazza, 2014). These variations may have an effect on whether ratio effects are observed. Agrillo et al. (2015) found that simultaneous presentation of dot arrays to adult humans led to ratio dependence in both the small and large number range, whereas sequential presentation of dot arrays led to ratio independence in the small number range. Further, in sequential arrays when distractors were intermingled there was ratio dependence in the small number range but no ratio effect when distractors were separated. Odic et al. (2014) showed that 5-year-old children's history of high or low confidence (manipulated via easy or difficult discriminations) and the presence or absence of feedback on preceding trials affects their ratio sensitivity on a numerical discrimination task.

Extensive training with feedback can improve numerical performance and have an effect on ratio sensitivity (Bisazza et al., 2014; DeWind and Brannon, 2012). Human adults given feedback in four sessions over a period of two weeks (648 trials per session) showed rapid improvement in their ability to compare numerosities using dot arrays (DeWind and Brannon, 2012). A ratio effect was present from the first to last session, but there was a reduction in the Weber fraction. Guppies trained for up to 120 trials for each discrimination (2 vs. 3, 3 vs. 4, 4 vs. 5, and 5 vs. 6) succeeded on ratios (0.75 and

0.80; Bisazza et al., 2014) that fish failed in other studies with less training (Agrillo et al., 2012a).

Number discrimination studies in mammals and birds that employ extensive training often include hundreds or thousands of training trials (e.g., Beran, 2007; Jaakkola et al., 2005; Pepperberg, 2006; Tomonaga, 2008; Yaman et al., 2012), whereas the only study of extensive training with fish subjects presented up to 120 training trials (Bisazza et al., 2014). Providing fish with a larger number of training trials, more similar to the studies of mammals and birds, may reveal better overall performance and inform the debate about whether one or two number representation systems operate in non-human animals. The aim of the present study was to present goldfish with relative quantity task using three ratios (0.33, 0.5, and 0.67) after extensive training of approximately 1200 trials provided across two blocks of training sessions. Thus, we increased the number of training trials by a magnitude of ten compared to a previous study on extensive training in fish (Bisazza et al., 2014). We presented stimuli in both the small number range (1 vs. 3, 2 vs. 4, 2 vs. 3) and the large number range (5 vs. 15, 6 vs. 12, 10 vs. 15) to investigate whether a ratio effect occurs in either number range after extensive training.

This study represents the first time the numerical ability of goldfish has been tested. The numerical competence of other fish species has been assessed, including guppies (e.g., Agrillo et al., 2014; Dadda et al., 2015), mosquitofish (e.g, Agrillo et al., 2007, 2011; Dadda et al., 2009), three-spined sticklebacks (Mehlis et al., 2015), angelfish (Agrillo et al., 2012b; Gomez-Laplaza and Gerlai, 2011a,b, 2013), Siamese fighting fish (Agrillo et al., 2012b), zebrafish (Agrillo et al., 2012b), redtail splitfin (Agrillo et al., 2012b), goldbelly top-minnows (Agrillo and Dadda, 2007), peacock blennies (Hennig, 1977) and mollies (Bradner and McRobert, 2001). Goldfish have been frequently tested in studies of visual perception and are notable as models for evaluating the visual system (e.g., Mora-Ferrer and Neumeyer, 2009; Wyzisk and Neumeyer, 2007).

2. Methods

2.1. Subjects

The subjects were four commercially-obtained goldfish (*Carassius auratus*), 8–11 cm in total length (N = 4). Individual fish were identified by their phenotypic characteristics (but sex could not be determined phenotypically). All fish were obtained in 2011 and were approximately three years old at the start of the study. Subjects were housed in pairs in 10 gallon tanks (50 cm long × 26 cm wide × 31 cm high). Each tank contained aquarium gravel and flat glass marbles, one artificial plant, and an Aqueon Quiet Flow 10 filter unit that provided aerated, filtered, and conditioned water. The top of each tank contained one Aqueon 10 W mini-compact fluorescent light bulb that simulated a natural light/dark cycle. The tanks were covered on three sides by blue cellophane on the exterior walls. Water temperature was kept between 22 °C–25 °C. Water changes (40%) were carried out on a weekly basis and tanks were monitored regularly for pH and waste levels with Tetra Aquarium Test Strips (Tetra U.S., Blacksburg, VA). Training and testing sessions were conducted during daylight hours (morning and/or afternoon). On days with training or test sessions subjects were fed only during the sessions. On weekdays when the fish were not trained or tested, they were fed twice: once with TetraFin flakes (Tetra GMBH, Melle, Germany) and once with API Premium pellets (Mars Fishcare Inc., Chalfont, PA).

Fish 4 was excluded from the experiment after failing to cooperate (would not approach stimulus board) during training block 1 prior to the first test block. As a result, three fish completed the first test block. Fish 3 developed a pervasive side bias during the second

Fig. 1. Examples of the experimental stimuli showing 6 of the 24 pairs. One example from the small number range (left column) and one example from the large number range (right column) is shown for each of the three ratios. The examples in the left column are controlled for density (inter-dot interval) and the examples in the right column are controlled for space (space encompassed by the most distant dots in the array). The examples for 0.33 are not controlled for surface area (summed area of dots), so the dots are the same size within a stimulus pair. The examples for 0.50 and 0.67 are controlled for surface area, so the dots are different sizes within a stimulus pair.

Fig. 2. Schematic diagram of the experimental set up. (A) Test tank showing the position of the stimulus board (dark gray shaded region). The fish had to be located within the light gray shaded region when the stimulus board was placed in the tank at the beginning of a session. (B) The stimulus board side used for training stage 1 (S+ only). (C) The stimulus board side used for training stage 2 and the testing blocks (S+ and S−).

training block in between the two test blocks and was excluded from the second test block, but fish 1 and fish 2 completed the entire experiment. Fish 1 and fish 2 were tested in a study on shape discrimination and a pilot study on quantity discrimination prior to the current study (DeLong et al., 2016; Keller and DeLong, 2014). Fish 3 and 4 were trained to eat from a syringe prior to this study but were otherwise experimentally naïve.

2.2. Stimuli

Fig. 1 shows examples of the stimulus pairs. The stimulus pairs consisted of arrays of black dots on a white background. The dots were 0.3 cm–0.5 cm in diameter. Three different ratios were used: 0.33, 0.5, 0.67. Stimuli representing the ratio 0.5 were presented during the training sessions and all three ratios were presented during the test sessions. Each of the three ratios were presented in a low number range (1–4 dots) and a high number range (5–15 dots) as follows: 0.5 (2 vs. 4 and 6 vs. 12), 0.67 (1 vs. 3 and 5 vs. 15) and 0.67 (2 vs. 3 and 10 vs. 15).

To ensure that the fish attended to the numerical value of the stimuli and not continuous quantities, the dot array stimuli were controlled for three features that co-vary with number: cumulative surface area (summed area of dots), the density of items (inter-dot distance), and space occupied by the arrays (space encompassed by the most distant dots in the array) as in other studies (Bisazza et al., 2014; Miletto Petrazzini et al., 2014; Gomez-Laplaza and Gerlai, 2013). Half of the pairs were controlled for surface area so that the cumulative area of each array within a stimulus pair could not be used to select the correct choice. Since overall space and density are inversely correlated half the stimuli were controlled for space and half were controlled for density. Stimuli were created in Adobe Illustrator v. CS6 (Adobe Systems Incorporated, 2014). To calculate surface area, we used the area calculator plugin for Adobe Illustrator v. CS6. The area of the black space was matched within stimulus pairs for the half of the stimuli controlled for surface area. To control density, a line segment was created for each stimulus pair and each
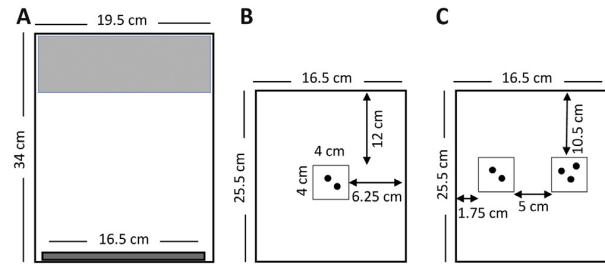
of the dots within each stimulus array were positioned at equal distances from each other using the segment. To control space, a line segment was created and the leftmost and rightmost two dots within each stimulus array were positioned at equal distances from each other using the segment. Aside from controlling placement of dots for space and density controls, the dots were randomly placed within the arrays.

In summary, for each of the six numerical contrasts (2 vs. 4, 6 vs. 12, 1 vs. 3, 5 vs. 15, 2 vs. 3, and 10 vs. 15) there were four different stimulus pairs (one pair controlled for surface area and density, one pair controlled for surface area and space, one pair controlled for density but not surface area, and one pair controlled for space but not surface area) for a total of 24 stimulus pairs. Thus, there were four different spatial arrangements of dots for each numerical contrast. Stimuli were printed using black ink on standard white paper and then laminated (as in experiment 5 in Siebeck et al., 2009). Each stimulus was centered on a 5 cm × 5 cm square (the stimulus card) with a 0.1 cm black border. The stimulus cards were attached to a white plastic corrugated stimulus board (25.5 cm high × 16.5 cm wide × 0.4 cm thick) with hook and loop fasteners.

2.3. Experimental set-up

Fig. 2 shows the experimental set-up. During training and testing sessions, subjects were moved to individual test tanks (3.7 gallons; 34.0 cm long × 19.5 cm wide × 21.5 cm high). Test tanks contained about 17 cm of conditioned water and no gravel or plants. Blue cellophane covered all four exterior side walls to block external stimuli during testing. There were three identical test tanks and fish rotated among the three tanks according to a random schedule. During training and testing sessions, the stimulus board was inserted in the front of the tank and rested against the interior tank wall. On one side of the stimulus board a single stimulus card could be presented that was centered on the board during training stage 1 (see Fig. 2B). On the other side of the stimulus board two stimulus cards could be presented during training stage 2 or the testing stage (see Fig. 2C).

2.4. Procedure

Sessions were conducted once or twice per day, always around the same time of day in the morning and afternoon (this resulted in a food deprivation period of about 3–4 h between sessions on the same day). There were typically 4–10 sessions per week. Test tanks were filled with water and water conditioner, then the fish was placed into the test tank using a small net and allowed to acclimate to the tank for a minimum of 15 min. Stimulus cards were placed on the hook and loop fasteners on the stimulus board. Before the stimulus board was placed in the tank the fish was ushered into the

back of the tank by gently tapping on the water in the back of the tank or tapping the top of the back wall of the tank (see Fig. 2A). This ensured the fish was swimming in the back of the tank when the stimulus board was placed in front of the tank at the beginning of a trial and encouraged the fish to examine both stimuli on the board before making a choice.

The fish were reinforced for correct choices with food paste; a mixture of TetraFin flakes and water (similar to Siebeck et al., 2009). Food paste was delivered to the fish using a 1.0 mL Luer-Lok tip syringe (BD, Franklin Lakes, NJ), and 0.01 mL of food paste was delivered for each correct choice. The fish had to tap (bite or push) the rewarded stimulus (S+) once before the food reward was given. The experimenter held the syringe out of sight from the fish behind the test tank wall until a correct choice was made. As soon as the subject tapped once on the rewarded stimulus, the experimenter lowered the syringe into the tank to the top center section of the S+ stimulus card. Once the syringe was positioned near the S+ stimulus card, the fish ate the food paste. If a fish tapped the incorrect stimulus (S−) it was not reinforced and the board was withdrawn. The correction method was not used, so the fish was allowed to choose only one stimulus (S+ or S−) for each trial. The stimulus board was always removed from the test tank during the intertrial interval of about 10–15 s when the experimenter recorded the data for the trial.

The fish's choice (S+ or S−) and the location of the choice (e.g., tap to the bottom, center, or top of the stimulus card) were recorded on each trial. A single experimenter presented the stimulus board, delivered the food reinforcement, and recorded the data during an experimental session. The experimenter stood in a position that allowed her to view the fish's behavior and deliver the food while holding the syringe out of view of the fish. Four experimenters trained and tested the fish throughout the experiment. The total session time was recorded for each session (rounded to the nearest minute), but individual trial times were not recorded. Sessions took approximately 3–5 min ($M = 4.0$ min for training sessions and $M = 3.3$ min for test sessions).

2.4.1. Training

Two of the fish were reinforced to select the low numerosity (fish 2 and fish 4) and the other two fish (fish 1 and fish 3) were trained to select the high numerosity in any given stimulus pair. For example, when presented with the pair 2 vs. 4, the low numerosity fish were reinforced when they chose two dots and the high numerosity fish were reinforced when they chose four dots. Before beginning training, the fish had to acclimate to the test tanks and feeding from the syringe during pre-training. Fig. 3A shows a fish eating from the syringe in the test tank during pre-training. All fish were pre-trained prior to this study for a minimum of 3–5 sessions (5 trials per session), although fish 4 was given 10 additional refresher pre-training sessions for this study because it was more hesitant to approach the syringe than the other fish.

In training stage 1, the fish had to tap the S+ to receive a food reward when only the positive stimulus but not the negative stimulus was presented (see Fig. 3B). The stimuli used in this stage were the S+ stimuli for the eight stimulus pairs for the 0.5 ratio. The low numerosity fish were presented with the two dot array for 2 vs. 4 and six dot array for 6 vs. 12 and the high numerosity fish were presented with the four dot array for 2 vs. 4 and twelve dot array for 6 vs. 12. There were 8 sessions, one session for each of the eight stimulus pairs for the 0.5 ratio. There were 5 trials per session. The fish readily tapped the S+ for all trials in the 8 sessions. Stage 1 was completed in five days (1–2 sessions per day).

In training stage 2, the fish were presented with the S+ and S- in a two-alternative forced choice task (see Fig. 3C). The stimuli used in this stage were the eight stimulus pairs for the 0.5 ratio. The position of the S+ for each trial was determined using a modified

pseudorandom Gellermann series (Gellermann, 1933) in which the S+ was never shown more than two trials in a row on the same side to prevent potential side biases. The fish were prompted with the correct answer under two conditions: (1) if they did not make a choice within three minutes of the stimulus board being placed into the test tank, or (2) if they made three incorrect choices in a row. A prompt consisted of presenting the syringe in front of the S+ at the beginning of a trial, so that the fish was forced to select the S+. Prompted trials were not included in the reported choice accuracy for training trials. The fish completed two blocks of training trials in stage 2–one prior to test block 1, and one prior to test block 2.

2.4.1.1. Training block 1.
Training block 1 began the same day training stage 1 was completed and lasted from June 30, 2014 to September 5, 2014. There were 64 total sessions (eight presentations of each of the eight stimulus pairs for the 0.5 ratio). The fish were required to complete 5 correct trials in a session, and there were 5–13 trials per session needed to achieve that goal. Fish 1 completed 452 total trials (5 were prompted trials), fish 2 completed 437 total trials (6 were prompted trials), and fish 3 completed 479 total trials (16 were prompted trials).

2.4.1.2. Training block 2.
Training block 2 began seven days after the completion of test block 1 and lasted from October 29, 2014 to June 22, 2015. There were 125 sessions (15–16 presentations of each of the eight stimulus pairs for the 0.5 ratio). There were always 6 trials per session for a total of 750 trials. All fish developed a pervasive side bias during the first 30–40 sessions (fish 1 and 2 had a left side bias, and fish 3 alternated between a right and left side bias). This might have been because the fish were trained at the unusually low rate of one or two times per week for the first 17 sessions due to availability of experimenters. Prompting (presenting the syringe with the food reward in front of the S+) was used to try to eliminate the side biases in sessions 28–83, and the number of training sessions was increased to a minimum of 4–5 sessions per week for the remaining 108 sessions. By session 83 the side biases had been eliminated, so sessions 84–125 contained no prompting.

2.4.2. Testing

Test sessions included both the S+ and S- in a two-alternative forced choice task like training stage 2 (see Fig. 3C). All 24 stimulus pairs were presented during test sessions, including the 8 stimulus pairs representing 0.5 used in the training stage and 16 novel stimulus pairs representing the 0.33 and 0.67 ratios. Each test session consisted of 6 trials. The first two trials had two of the eight stimulus pairs representing 0.5 (the S+ was on the right for one trial and the left for one trial). The last four trials in a session had four of the novel testing stimulus pairs representing 0.33 or 0.67. As in training stage 2, the position of the S+ for each trial was determined by sequences based on a modified pseudorandom Gellermann series with no more than two appearances of S+ on the same side. The S+ was on each side equally often within a test session (3 times on the right, 3 times on the left). The stimulus pairs selected for each test session were determined randomly, with the constraint that no stimulus pair was allowed to appear more than twice per session (thus there were 5 or 6 different stimulus pairs per session). The S+ for each stimulus pair was presented equally often on the right and the left across each block of test sessions. There were two blocks of test sessions.

2.4.2.1. Test block 1.
Test block 1 began ten days after the completion of training block 1 and lasted from September 15, 2014 to October 22, 2014. Test block 1 contained 24 test sessions, in which each of the 24 stimulus pairs was presented 6 times.
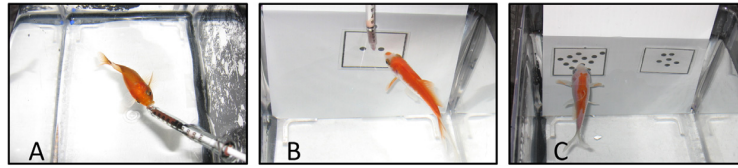
Fig. 3. Photographs of fish engaged in training and testing. (A) A fish eating flake paste from the syringe during the pretraining phase. (B) A fish being reinforced with flake paste from the syringe just after tapping the S+ during training stage 1. (C) A fish during training stage 2 approaching the S+. After tapping the S+, the syringe was inserted from above and positioned near the fish's head. If the S− was tapped the board was removed. The set-up was the same for training stage 2 and the test sessions (both S+ and S− presented simultaneously during a session).

2.4.2.2. Test block 2. Test block 2 began on the same day training block 2 was completed and lasted from June 22, 2015 to July 23, 2015. The 24 test sessions of test block 2 were exactly the same as test block 1 in terms of the composition of the sessions (the stimuli were presented in the same order).

2.5. Data analysis

Statistical analyses were performed using R (R Core Team, 2016). An experimental type I error level of $\alpha = .05$ was used. All p-values were adjusted using Hommel's (1988) method to control the experimental error level at $\alpha = .05$. In the case of post hoc comparisons, p-values were adjusted using Scheffé's (1959) method to control the experimental error at $\alpha = .05$. For the three models presented below, logistic regression was used to determine if treatment factors (stimulus ratio, number range: small vs. large, controlling stimulus surface area, controlling stimulus density vs. space, S+ position: right or left), differences between blocks of similar groups (fish, experimental block, test tank, day of week, morning vs. afternoon session), or covariates (session number, time of day) affected discriminative accuracy as measured by a choice to the S+ (correct choice) or to the S− (incorrect choice).

Model goodness-of-fit was assessed using residual deviance, the Hosmer-Lemeshow test (Hosmer and Lemeshow, 1980), and Bayesian Information Criterion (BIC; Schwarz, 1978) where appropriate. Variable selection was guided by bidirectional stepwise elimination during which the criterion for variable inclusion or exclusion was determined by the minimization of BIC (Venables and Ripley, 2002). Tests of overall accuracy against chance (50%) performance in a training or testing block were performed using the predicted marginal probability of a correct choice to the S+ from the logistic regression model at an average session level (e.g., session 31.1 of 62 sessions).

3. Results

3.1. Training sessions

3.1.1. Training block 1

In the first training block, the overall accuracy after controlling for session number was significantly better than chance (i.e., 50%), *Accuracy* = 70.9%, *SE* = 1.3%, *Z* = 14.1, *p* < 0.00001. The optimal logistic regression model was a significant predictor of accuracy as suggested by a Hosmer-Lemeshow test of model goodness-of-fit, $\chi^2(8) = 13.0$, $p = 0.1105$. The main effect of session and the interaction between fish and S+ location were both significant after Hommel's (1988) adjustments for multiple comparisons at $\alpha < 00.0001$. Sequential analysis of deviance using likelihood ratio tests was performed to test the significance of the main effects and the interaction effects. No other treatment factors, blocks, covariates, or associated interactions were statistically significant. Accuracy significantly improved as a function of the number of sessions, $\chi^2(1) = 27.1$, $p < 0.0001$. Post hoc analyses for the session main effect revealed that accuracy increased significantly over time
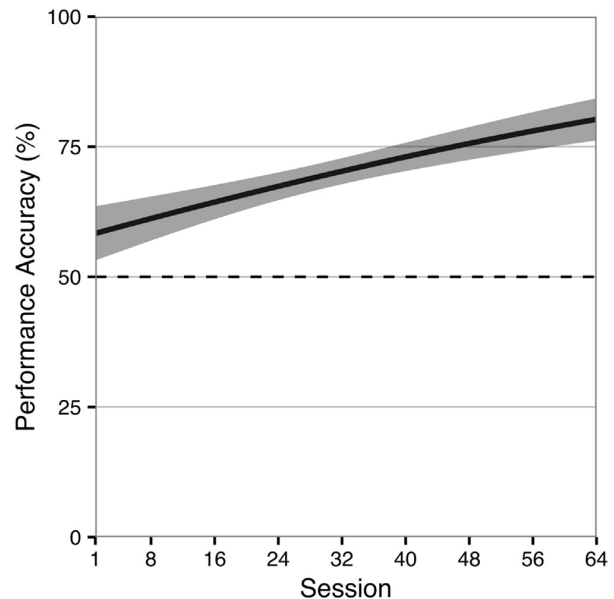

Fig. 4. The performance accuracy of the fish in training block 1. Performance significantly increased over time. Performance was significantly above chance for all sessions (model estimated accuracies are shown along with 95% simultaneous confidence intervals).

as shown in Fig. 4. Performance was always significantly better than chance for all sessions.

The main effect of fish was not significant, $\chi^2(2) = 5.9$, $p = 0.0172$, and the main effect of S+ location was not significant, $\chi^2(1) = 0.6$, $p = 0.4363$, but there was a significant interaction between the fish and S+ location, $\chi^2(2) = 456.0$, $p < 0.00001$. Fish 1 demonstrated a side bias, performing significantly better when the S+ was on the left (80.3%, *95% CI:* (74.4%, 85.1%)) than on the right (62.6%, *95% CI:* (56.1%, 68.6%)). Fish 2 also demonstrated a side bias, performing significantly better when the S+ was on the right (84.1%, *95% CI:* (78.8%, 88.3%)) than on the left (61.0%, *95% CI:* (54.1%, 67.6%)). Fish 3 did not exhibit a side bias; accuracy when the S+ was on the left side (65.3%, *95% CI:* (58.7%, 71.3%)) was not significantly different than when the S+ was on the right side (66.6%, *95% CI:* (60.4%, 72.3%)). Despite the side biases, all three fish performed significantly better than chance for both S+ locations as clearly shown by the simultaneous 95% confidence intervals exclusion of 50% accuracy from their regions.

The performance accuracy of the fish for all stimulus pairs in training block 1 is shown in Table 1 along with statistical results described below. There was no significant difference in performance for the small vs. large number range, and accuracy was not statistically significantly impacted by a stimulus pair's surface area, density, or space attributes. All three fish performed significantly better than chance on all eight stimulus pairs except for Fish 3 on

Table 1
Performance Accuracy in Training Blocks.

Pair #	Stimulus	Training Block 1			Training Block 2		
		Accuracy (%)	Z Test	Hommel's P-Value	Accuracy (%)	Z Test	Hommel's P-Value
1	2 vs.4 SA = D =	67.3	4.52	<0.0001	78.3	4.38	0.0001
2	2 vs.4 SA = S =	73.8	6.02	<0.0001	79.2	4.96	<0.0001
3	2 vs.4 SA ≠ D =	69.6	5.08	<0.0001	76.7	4.14	0.0002
4	2 vs.4 SA ≠ S =	68.2	4.75	<0.0001	75.0	3.87	0.0003
ALL	2 vs.4	69.7	10.19	<0.0001	77.4	8.70	<0.0001
5	6 vs.12 SA = D =	72.8	5.80	<0.0001	72.7	3.69	0.0004
6	6 vs.12 SA = S=	59.0	2.47	0.0135	75.9	3.81	0.0003
7	6 vs.12 SA ≠ D =	73.8	6.02	<0.0001	72.2	3.77	0.0003
8	6 vs.12 SA ≠ S =	74.1	6.13	<0.0001	71.7	3.36	0.0008
ALL	6 vs.12	69.5	10.11	<0.0001	73.0	7.30	<0.0001

Note. Performance accuracy was statistically significantly greater than chance (50%) for all stimulus pairs using Hommel's method for multiple comparisons at an overall $\alpha = 0.05$. SA = indicates surface area was controlled, SA ≠ indicates surface area not controlled, D = indicates density was controlled, S = indicates space was controlled.

stimulus pair 1 with a 2 vs. 4 ratio controlled for space and density (*Accuracy* = 56.9%, *SE* = 6.1%, *Z* = 1.1, *p* = 0.2596) and on stimulus pair 6 with a 6 vs. 12 ratio controlled for surface area and space (*Accuracy* = 46.6%, *SE* = 5.8%, *Z* = −0.6, *p* = 0.7213).

3.1.2. Training block 2

Since many of the early training block 2 sessions contained prompted trials for correcting a side bias in the fish which was eliminated by session 83, we chose to model only sessions 84–125 (none of these 42 sessions contained any prompted trials). To fairly compare performance between the first and second training blocks, we compared the final 42 sessions in both training blocks. The log-odds in the second training block after controlling for session number (i.e., the accuracy at the mid-point, session 104.5) was compared to the log-odds in the last 42 sessions of the first block after controlling for session number (i.e., the accuracy at the mid-point of the last 42 sessions, session 43.5). Accuracy in the final 42 sessions of the second training block (78.6%) was not significantly different from accuracy in the final 42 sessions of the first training block (75.2%), *Z* = 1.2, *p* = 0.0969.

In the second training block, the overall accuracy after controlling for session number was significantly better than chance (i.e., 50%), *Accuracy* = 78.6%, *SE* = 2.1%, *Z* = 10.4, *p* < 0.00001. The optimal logistic regression model was a significant predictor of accuracy as suggested by a deviance test of model goodness-of-fit, $\chi^2(500) = 513.8$, $p = 0.3248$. A Hosmer-Lemeshow test of model goodness-of-fit further supported use of the logistic regression model, $\chi^2(8) = 6.7$, $p = 0.5671.$.

The S+ location main effect and the interaction between session and S+ location were both significant after Hommel's (1988) adjustments for multiple comparisons at $\alpha < 0.05$. Sequential analysis of deviance using likelihood ratio tests was performed to test the significance of the main effects and the interaction: a main effect of session number was not significant, $\chi^2(1) = 2.6$, $p = 0.1064$, but the main effect of S+ location was significant, $\chi^2(1) = 39.9$, $p < 0.00001$. There was a significant interaction between the session number and S+ location, $\chi^2(1) = 8.3$, $p = 0.0081$. No other treatment factors, blocks, covariates, or associated interactions were statistically significant.

As seen in Fig. 5, accuracy when the S+ was on the right did not change significantly over time while accuracy when the S+ was on the left increased significantly over time: At the beginning of the second training block, there was no significant evidence of side bias (*Session* = 84, *Left Side Accuracy* = 72.2%, *95% CI:* (58.3%, 82.8%)), *Right Side Accuracy* = 64.3%, *95% CI:* (51.9%, 75.0%). By the middle of the second training block, evidence of a significant side bias emerged; accuracy was significantly better when the S+ was on the left side (89.0%, *95% CI:* (84.1%, 92.6%)) than on the right side (63.2%, *95% CI:* (57.0%, 68.9%)). By the end of the second training block, the side bias
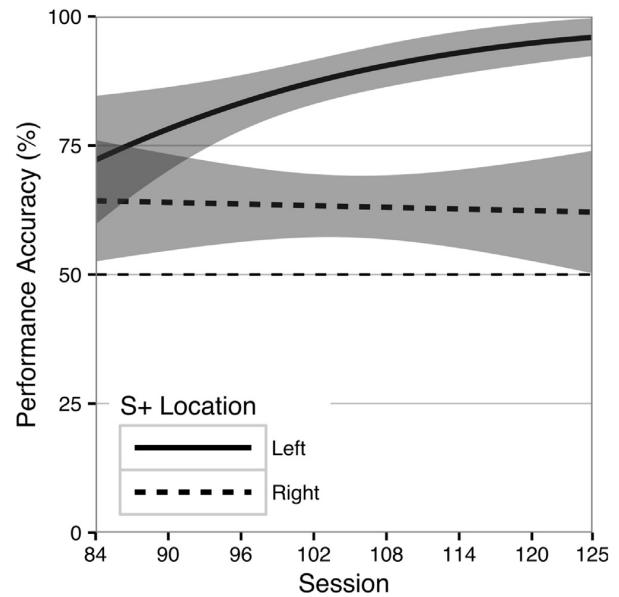

Fig. 5. The performance accuracy of the fish in training block 2 for sessions 84–125. Performance significantly increased over time when the S+ was on the left, but did not change over time when the S+ was on the right. Performance was significantly above chance for all sessions (model estimated accuracies are shown along with simultaneous 95% confidence intervals).

became more pronounced; accuracy when the S+ was on the left side (96.0%, *95% CI:* (90.2%, 98.4%)) was significantly greater than on the right side (62.1%, *95% CI:* (49.7%, 73.1%)) and significantly greater than accuracy for the left side in the middle of the training block (89.0%, *95% CI:* (84.1%, 92.6%)).

The performance accuracy of the fish for all stimulus pairs in training block 2 and the statistical analyses are presented in Table 1. There was no significant difference in performance for the small vs. large number range, and accuracy was not statistically significantly impacted by a stimulus pair's surface area, density, or space attributes. Both fish performed significantly better than chance on all stimulus pairs.

3.2. Testing sessions

3.2.1. Performance accuracy for blocks 1 and 2

As a group, the fish successfully discriminated between the stimuli pairs across all stimuli pairs. The overall accuracy was significantly better than chance in the first test block (i.e., 50%), *Accuracy* = 91.0%, *SE* = 1.8%, *Z* = 10.5, *p* < 0.00001. Accuracy in

Table 2
Post Hoc Contrasts for the Interaction Between Test Block and Stimulus Ratio.

Stimulus Ratio	Test Block	Accuracy (%)	95% CI (%)	Grouping
0.33	1	84.0	(77.1, 89.1)	AB
	2	94.8	(88.1, 97.8)	BC
0.50	1	97.9	(93.7, 99.3)	C
	2	92.7	(85.5, 96.5)	ABC
0.67	1	80.6	(73.3, 86.2)	A
	2	96.9	(90.8, 99.0)	BC

Note. For each experimental block and stimulus ratio, the probability of choosing the S+ is given with corresponding Scheffé-adjusted 95% confidence intervals. All probabilities are significantly greater than chance (i.e., 50%). Block-ratio combinations are grouped: probabilities for block-ratio pair combinations that share a group were not statistically significantly different.

the second test block was also significantly better than chance (*Accuracy* = 95.1%, *SE* = 1.4%, *Z* = 10.5, *p* < 0.00001). Accuracy in the first test block was significantly better than accuracy in the first training block (70.9%, *Z* = 6.2, *p* < 0.00001). Furthermore, accuracy in the second test block (95.1%) was significantly higher than accuracy in the second training block (78.6%, *Z* = 5.4, *p* < 0.00001).

The optimal logistic regression model for the two test blocks was a significant predictor of the probability of fish choosing the S+ as suggested by a non-significant deviance-based goodness-of-fit test, $\chi^2(714) = 413.6$, $p = 1$, and a non-significant Hosmer-Lemeshow test, $\chi^2(3) = 0$, $p = 1$. The model retained one second-order interaction and its corresponding main effects, all of which were significant after Hommel's (1988) adjustments for multiple comparisons at $\alpha < 0.001$. Sequential analysis of deviance using likelihood ratio tests was performed to test the significance of the main effects and the interaction: accuracy was significantly impacted by test block, $\chi^2(1) = 11.41$, $p = 0.0008$, the stimulus ratio, $\chi^2(2) = 14.23$, $p = 0.0008$, and the interaction between the test block and stimulus ratio, $\chi^2(2) = 15.51$, $p = 0.0008$. No other treatment factors, blocks, or covariates, or associated interactions were significant.

Post hoc analyses given in Table 2 for the interaction between experimental block and stimulus ratio revealed that accuracy was best in the first test block of the experiment for the 0.5 ratio. There was no difference in accuracy in the first block between the 0.33 and the 0.67 ratios. There were no differences in accuracy in the second block among any of the stimulus ratios. Accuracy significantly improved from the first block to the second block for the 0.67 ratio, but there was no change in accuracy from the first block to the second block for the 0.33 and 0.5 ratios. Accuracy in both experimental blocks for all three stimulus ratios was always significantly greater than chance as shown in Fig. 6.

The performance accuracy of the fish for all 24 stimulus pairs in both testing blocks are shown in Table 3. As a group, the fish performed significantly better than chance on all stimulus pairs. There were 24 stimuli test block 1 and 24 stimuli in test block 2 for 48 stimuli-block comparisons – the fish performed significantly better than chance in 42 of these (88%), however in four of these cases significance was marginal (*p* = 0.0548, M = 77.8%). Both cases in which the fish's performance was not different from chance (*p* < 0.05) occurred in test block 1; one case for the numerical contrast 1 vs. 3 and one case for the numerical contrast 10 vs. 15.

Table 4 provides a summary of performance accuracy across the study as a function of controlled continuous quantities and number range. Across all training and testing blocks, there was no significant difference in performance between pairs in the small vs. large number range, and accuracy was not statistically significantly impacted by a stimulus pair's surface area, density, or space attributes.
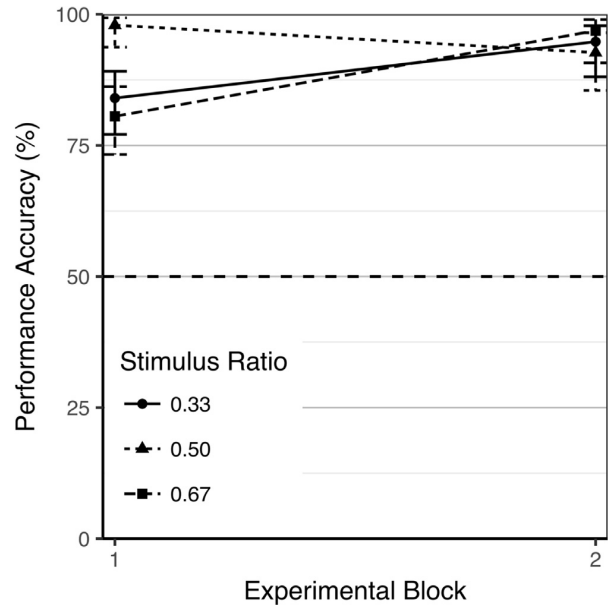

Fig. 6. The performance accuracy of the fish in test blocks 1 and 2. In test block 1, performance accuracy was significantly higher on 0.50 compared to the other ratios, and did not differ between 0.33 and 0.67. In test block 2, there was no significant difference among the three ratios. Performance was significantly above chance for all ratios in both blocks (model estimated accuracies are shown along with simultaneous 95% confidence intervals).

3.2.2. First trial performance

The performance of each fish on first exposure to each of the testing stimuli in test block 1 is shown in Table 5. First trial performance shows the ability of the fish to discriminate between stimuli before any feedback was given and learning could occur. Overall, the fish were correct on 79.2% of first trials, which is only about 10% lower than their overall performance on test block 1 (90.9%). There was no difference in first trial performance when comparing the ratios 0.33 and 0.67, although fish 3 got only 50% correct on 0.67 and 100% correct on 0.33. Two or three fish had correct choices on each stimulus pair with the exception of two pairs (one pair for 0.33 and one pair for 0.67). First trial performance did not seem to be affected by whether the stimuli came from the small vs. large number range or a stimulus pair's surface area, density, or space attributes. However, individual differences between fish in first trial performance as well as performance on training block 1 may indicate differences in learning strategies, cognitive style, or cue preference (Agrillo et al., 2009; Lucon-Xiccato and Bisazza, 2016; Mesquita et al., 2015).

4. Discussion

Previous studies show enhanced performance on numerical tasks when subjects are given extensive training (e.g., Agrillo and Bisazza, 2014; DeWind and Brannon, 2012). Humans, other mammals, and birds have received hundreds or thousands of trials in extensive training procedures, whereas fish have typically been given 10–30 trials (Agrillo et al., 2012a) or 120 trials at most (Bisazza et al., 2014). The purpose of the present study was to investigate the performance of goldfish given approximately 1200 training trials, which exceeds the amount of training given to any fish in previous numerical discrimination tasks. The fish achieved 91% correct after about 450 training trials and 95% correct after an additional 750 training trials (there were about 300 test trials). This overall level of performance is higher than previous work with fish that received

Table 3
Performance Accuracy in Test Blocks.

Pair #	Stimulus	Test Block 1			Test Block 2		
		Accuracy (%)	Z Test	Hommel's P-Value	Accuracy (%)	Z Test	Hommel's P-Value
1	2 vs.4 SA = D =	94.4	3.77	0.0026	100.0	3.46	0.0049
2	2 vs.4 SA = S =	100.0	4.24	0.0004	91.7	2.89	0.0116
3	2 vs.4 SA ≠ D =	100.0	4.24	0.0004	91.7	2.89	0.0116
4	2 vs.4 SA ≠ S =	100.0	4.24	0.0004	83.3	2.31	0.0209
ALL	2 vs.4	98.6	8.25	<0.0001	91.7	5.78	<0.0001
5	6 vs.12 SA = D =	100.0	4.24	0.0004	91.7	2.89	0.0116
6	6 vs.12 SA = S=	100.0	4.24	0.0004	100.0	3.46	0.0049
7	6 vs.12 SA ≠ D =	94.4	3.77	0.0026	91.7	2.89	0.0116
8	6 vs.12 SA ≠ S =	94.4	3.77	0.0026	91.7	2.89	0.0116
ALL	6 vs.12	97.2	8.01	<0.0001	93.8	6.07	<0.0001
9	1 vs.3 SA = D =	83.3	2.83	0.0274	100.0	3.46	0.0049
10	1 vs.3 SA = S =	94.4	3.77	0.0026	100.0	3.46	0.0049
11	1 vs.3 SA ≠ D =	88.9	3.30	0.0102	83.3	2.31	0.0209
12	1 vs.3 SA ≠ S =	72.2	1.88	0.1202	91.7	2.89	0.0116
ALL	1 vs.3	84.7	5.89	<0.0001	93.8	6.07	<0.0001
13	5 vs.15 SA = D =	83.3	2.83	0.0274	100.0	3.46	0.0049
14	5 vs.15 SA = S =	88.9	3.30	0.0102	100.0	3.46	0.0049
15	5 vs.15 SA ≠ D =	83.3	2.83	0.0274	91.7	2.89	0.0116
16	5 vs.15 SA ≠ S =	77.8	2.36	0.0548	91.7	2.89	0.0116
ALL	5 vs.15	83.3	5.65	<0.0001	95.8	6.35	<0.0001
17	2 vs.3 SA = D =	77.8	2.36	0.0548	100.0	3.46	0.0049
18	2 vs.3 SA = S =	77.8	2.36	0.0548	100.0	3.46	0.0049
19	2 vs.3 SA ≠ D =	77.8	2.36	0.0548	100.0	3.46	0.0049
20	2 vs.3 SA ≠ S =	83.3	2.83	0.0274	91.7	2.89	0.0116
ALL	2 vs.3	79.2	4.96	<0.0001	97.9	6.64	<0.0001
21	10 vs.15 SA = D =	66.7	1.42	0.1556	91.7	2.89	0.0116
22	10 vs.15 SA = S =	83.3	2.83	0.0274	100.0	3.46	0.0049
23	10 vs.15 SA ≠ D =	88.9	3.30	0.0102	91.7	2.89	0.0116
24	10 vs.15 SA ≠ S =	88.9	3.30	0.0102	100.0	3.46	0.0049
ALL	10 vs.15	81.9	5.41	<0.0001	95.8	6.35	<0.0001

Note. The eight stimulus pairs for the ratio 0.5 were used in the training phases. All 24 stimuli were used in the two test blocks. All accuracies were statistically significantly greater than chance (50%) for all stimulus pairs using Hommel's method for multiple comparisons at an overall $\alpha = 0.05$. SA = indicates surface area was controlled, SA ≠ indicates surface area not controlled, D = indicates density was controlled, S = indicates space was controlled.

Table 4
Performance Accuracy (% correct) for Both Training and Test Blocks.

Variable	Training Block 1	Test Block 1	Training Block 2	Test Block 2	Overall (Weighted)
Surface Area Controlled	67.8	87.5	76.6	97.9	76.2
Surface Area Not Controlled	71.4	87.5	73.8	91.7	76.9
Density Controlled	70.8	86.6	74.8	94.4	76.9
Space Controlled	68.4	88.4	75.6	95.1	76.2
Small Number Range	69.7	87.5	77.4	95.1	77.1
Large Number Range	69.5	87.5	73.0	94.4	76.0

Table 5
First Trial Performance during Test Block 1.

Stimulus Pair	Fish 1	Fish 2	Fish 3	Overall
1 vs.3 SA = D =	1	1	1	100%
1 vs.3 SA = S =	1	1	1	100%
1 vs.3 SA ≠ D =	1	1	1	100%
1 vs.3 SA ≠ S =	0	0	1	33.3%
5 vs.15 SA = D =	1	0	1	66.7%
5 vs.15 SA = S =	1	1	1	100%
5 vs.15 SA ≠ D =	1	0	1	66.7%
5 vs.15 SA ≠ S =	0	1	1	66.7%
ALL 0.33	75%	62.5%	100%	79.2%
2 vs.3 SA = D =	1	1	1	100%
2 vs.3 SA = S =	1	1	0	66.7%
2 vs.3 SA ≠ D =	1	1	0	66.7%
2 vs.3 SA ≠ S =	1	1	0	66.7%
5 vs.15 SA = D =	0	1	0	33.3%
5 vs.15 SA = S =	1	1	1	100%
5 vs.15 SA ≠ D =	1	1	1	100%
5 vs.15 SA ≠ S =	1	1	1	100%
ALL 0.67	87.5%	100%	50%	79.2%
Overall	81.3%	81.3%	75%	79.2%

Note. The success of each fish on first exposure (1 = correct choice, 0 = incorrect choice) to a stimulus pair is given for test block 1. SA = indicates surface area was controlled, SA ≠ indicates surface area not controlled, D = indicates density was controlled, S = indicates space was controlled.

fewer training trials. For example, mosquitofish trained with less than 20 trials achieved about 55–65% correct on the ratios 1 vs. 2 and 2 vs. 3 (Agrillo et al., 2012a). Guppies trained with fewer than 20 trials achieved about 65–75% accuracy on moving and static stimuli for 1 vs. 2 and 80% accuracy on moving stimuli stimuli for 3 vs. 4 (Agrillo et al., 2014). Guppies receiving 24–120 training trials achieved about 65–75% correct on the ratios 2 vs. 3, 3 vs. 4, and 4 vs. 5 (Bisazza et al., 2014).

The performance accuracy of the fish in the present study is more comparable to studies with non-human primates than past studies with fish. Chimpanzees with 4000–8000 training trials attain accuracies up to 97%–100% (Tomonaga, 2008) and trained monkeys' performance ranged from 90 to 100% (Beran, 2008). Fish provided with the similar training opportunities as mammals and birds appear to be able to match their high accuracy levels. Extensive training may also reveal that fish can succeed on more difficult ratio discriminations (e.g., 0.83 or 0.88). The present study was not designed to test the upper limit of discrimination capabilities in fish – the highest ratio we tested was 0.67. However, Bisazza et al. (2014) found that providing up to 120 training trials allowed guppies to succeed on the ratios 3 vs. 4 and 4 vs. 5, whereas mosquitofish trained with fewer than 20 trials did not perform well on the same ratios (Agrillo et al., 2012a). This suggests researchers seeking to compare the upper limits of numerical skills of fish with birds and mammals adopt training procedures with more training trials (>1000).

We presented stimuli in both the small and large number range to examine whether one or two numerical representation systems (OTS and ANS) appeared to function during the task. Studies thus far on non-human animals have been equivocal, some supporting the idea of one system (e.g., Cantlon and Brannon, 2006, 2007; Jaakkola et al., 2005; Mehlis et al., 2015; Tomonaga, 2008) and others advocating for two systems (e.g., Agrillo et al., 2014, 2015; Bisazza et al., 2010; Piffer et al., 2012). The current results showed no significant difference in performance between stimuli in the small number range (1 vs. 3, 2 vs. 4, 2 vs. 3) and the large number range (5 vs. 15, 6 vs. 12, 10 vs. 15). Further, the performance of the fish after extensive training did not follow Weber's Law in either the small or large number range. In the first test block after about 450 training trials, the fish performed best on the ratio used during training (0.5) but there was no difference in performance between the two novel test ratios (0.33 and 0.67). That result can likely be attributed to the fishes' extensive training experience with the ratio 0.5. In the second test block (after an additional 750 trials) there was no difference in performance among all three ratios – there was no ratio effect.

Finding a lack of ratio sensitivity in either the small or the large number range is an unusual result found in few studies (Irie-Sugimoto et al., 2009; Irie and Hasegawa, 2012). Asian elephants tested in a spontaneous choice procedure with no training showed ratio insensitivity in the small (Irie-Sugimoto et al., 2009) and large number range (Irie and Hasegawa, 2012). However, elephants tested in another study with a spontaneous choice procedure showed a ratio effect in both number ranges (Perdue et al., 2012). Many studies find a ratio effect across both number ranges (Cantlon and Brannon, 2006, 2007; Cantlon et al., 2015; Jaakkola et al., 2005; Mehlis et al., 2015; Tomonaga, 2008). Consistency in performance across both number ranges in terms of ratio sensitivity has been suggested as evidence that one numerical system, the ANS, has been employed during a numerical task. One possibility is that only one numerical system exists for non-human animals (the ANS), but another possibility is that both the OTS and ANS are present but only the ANS has been recruited for the task at hand (i.e., a subitizing response has not been triggered). It has been suggested that the ANS can be recruited to represent numbers in the small number range in certain contexts (Bisazza et al., 2010; Cordes

and Brannon, 2008). In the present study half of the training stimuli pairs represented the large number range (6 vs. 12) and half represented the small number range (2 vs. 4). The repeated exposure to the training stimuli from the large number range (hundreds of trials) may have elicited the ANS to control responses within the context of the task during testing. The amount of exposure an individual has in general to numerical matching tasks can affect their performance and may influence which system controls their behavior. Number-experienced monkeys showed some differences in ratio dependency relative to number-naïve monkeys (Cantlon and Brannon, 2007).

An alternative explanation as to why these results were ratio-independent concerns the ratio comparisons utilized in this study. We tested ratios up to 0.67, and first trial performance of the fish suggests that 0.67 was not a difficult comparison for them ($M = 79.2\%$). It's possible that a ratio effect may have been observed if larger ratios were employed (e.g., 0.80 or 0.90). Extensive training may cause a ceiling effect in performance up to a point at which the upper limit of discrimination ability is approached. Non-human primates given extensive training on a relative quantity judgment task show performance at 90% or better, even on ratios up to 0.90, but a strong ratio effect was observed (Beran, 2008). The monkeys achieved 95–100% correct up to a ratio of 0.67, at which point performance decreased and became more variable (Beran, 2008). Future studies utilizing extensive training can explore this idea by testing higher ratios beyond 0.67.

Extensive experience over hundreds of trials may lead to the recruitment of additional representational systems (aside from the ANS and OTS) not typically involved in relative quantity judgments presented via spontaneous choice procedures or brief training regimes (Agrillo and Bisazza, 2014). In humans, rapid relative quantity judgments are processed in the intraparietal sulcus, but symbolic numerical abilities developed with a great deal of training involve other areas such as the prefrontal cortex, the insula, and the cerebellum (Arsalidou and Taylor, 2011). We do not claim that the fish developed advanced numerical abilities similar to humans nor that fish brains have equivalent brain structures to humans, but the general idea that performance and neural resources utilized after extensive training differ from those found after no training or brief training may hold true for non-human animals.

In this study, extensive training was confined to a single ratio (0.5), which is a ratio that many species can discriminate (e.g., *chicks* – Rugani et al., 2008; *jungle crows* – Bogale et al., 2014; *frogs* – Stancher et al., 2015; *salamanders* – Krusche et al., 2010). Other studies using training procedures have chosen the same 0.5 ratio for training (Agrillo et al., 2012a, 2014). It is interesting that the performance of the fish in the current study significantly improved on the ratio 0.67 in the second test block (2 vs. 3 and 10 vs. 15), even though the training trials focused exclusively on the ratio 0.5 (2 vs. 4 and 6 vs. 12). One potential explanation concerning stimuli in the small number range only (2 vs. 4 and 2 vs. 3) is that the fish trained to "choose low" learned to select the stimulus with two dots in both cases and the fish trained to "choose high" learned to avoid the stimulus with two dots in both cases. An alternative explanation is that a large number of training trials with any one ratio comparison eventually leads to an increase in performance across a range of ratio comparisons, but this idea would have to be tested in a subsequent study. Animals trained extensively on one ratio could be compared to those trained on two or more ratios, and both groups could be tested on novel ratios.

There are challenges and advantages to conducting a long-term training study with fish as there are with any non-human animal. The current study spanned 13 months, which is longer than a typical numerosity study conducted with fish subjects spanning days to weeks (e.g., Bisazza et al., 2014; Miletto Petrazzini et al., 2012)

and closer to the length of time dedicated to studies of mammals or birds that can span many months (Jakkola et al., 2005; Pepperberg, 2006). One limitation of these long-term studies tends to be a lower number of subjects (typically 1–10 individuals), a limitation shared by the present study. Studies of extensive training in fish must employ species that are robust in captivity for several years, which is why we selected goldfish as subjects (www.fishbase.org). Even with species with a long life span some individuals cease to respond to the task over the course of a long-term study, which further reduces the number of subjects. Collecting hundreds of trials of data from each individual compensated for having few subjects. However, we must still use caution in interpreting results from few subjects because within-species differences between fish can occur in studies of numerical discrimination (Lucon-Xiccato et al., 2016; Miletto Petrazzini and Agrillo, 2016). Since using few subjects can potentially bias the assessment of numerical abilities, it is important to replicate these results with additional subjects in future research.

An advantage to conducting an extensive training study is statistical power. The number of trials used in the present study afforded us very high power for each logistic regression model, enabling us to detect effects with far more precision and very low Type I and Type II error rates. For example, the pre-variable selection logistic regression model for training block 1 had power of 1.000 for medium (Cohen's $f^2 = 0.15$) and large effect sizes (Cohen's $f^2 = 0.35$, $\alpha = 0.05$, $df_1 = 53$, $df_2 = 1287$). Similarly, the pre-variable selection logistic regression model for training block 2 had power of 1.000 for medium and large effect sizes ($\alpha = 0.05$, $df_1 = 44$, $df_2 = 459$). For the test blocks, the initial logistic regression model before variable selection had power of 1.000 for medium and large effect sizes ($\alpha = 0.05$, $df_1 = 75$, $df_2 = 644$). If we had tried to estimate the impact of the same set of variables and interactions for either training block with only 100 trials, power to detect large effect sizes would have been 0.5590 ($\beta = 0.4410$) while power to detect medium effect sizes would have been 0.2175 ($\beta = 0.7825$). High power (and therefore low Type II error rates) is a crucial component of any experimental design and is often lacking in psychological studies, particularly when it is difficult to test a large number of subjects.

5. Conclusions

The goldfishs' performance exceeded 90% correct on the ratios 0.33, 0.5, and 0.67, which is higher than all previous studies in which fish were tested in training paradigms (Agrillo et al., 2012a, 2014; Bisazza et al., 2014). At the conclusion of training the performance of the fish was ratio-insensitive across both the small and large number range, which is an uncommon result. This may indicate that a single number representation system was at work, additional neuro-cognitive systems were engaged during the task as a result of the extensive training, or that the ratio comparisons we utilized (up to 0.67) were not large enough to reveal ratio dependency. Future research must include larger ratio comparisons (e.g., 0.75, 0.90) to avoid a potential ceiling effect.

This is the first study to provide evidence for numerical abilities in goldfish. Most of the research on numerical abilities in fish has been conducted with guppies (e.g., Agrillo et al., 2014; Dadda et al., 2015), mosquitofish (e.g, Agrillo et al., 2007, 2011; Dadda et al., 2009), and angelfish (Gomez-Laplaza and Gerlai, 2011a,b, 2013). These species are distantly related to goldfish. Angelfish (*Pterophyllum scalare*) are in the Cichlidae family, mosquitofish (*Gambusia holbrooki*) and guppies (*Poecilia reticulate*) are in the Poeciliidae family, and goldfish (*Carassius auratus*) are in the Cyprinidae family. The existence of numerical abilities in these disparate species

suggests that all teleost fish may possess quantification systems (Agrillo et al., 2012b).

In summary, these results suggest that goldfish given extensive training of over 1000 trials can achieve accuracy levels on a numerical discrimination task comparable to well-trained birds or non-human primates (e.g., Beran, 2008; Pepperberg, 2006). This is not surprising given that fish perceptual and cognitive abilities compare very well to other vertebrates on most tasks (Brown, 2015). Fish engage in complex cognitive tasks such as pattern recognition and categorization, short and long-term memory formation, tool use, and navigation using cognitive maps (Aronson, 1956; Brown et al., 2011; McAroe et al., 2016; Lopez et al., 1999). Like other non-human animals, fish likely use numerical information to maximize their fitness during foraging, predator defense, or threat assessment.

Ethical standards and conflict of interest

Acknowledgments

References

Agrillo, C., Bisazza, A., 2014. Spontaneous versus trained numerical abilities: a comparison between the two main tools to study numerical competence in non-human animals. J. Neurosci. Methods 234, 82–91.

Agrillo, C., Dadda, M., 2007. Discrimination of the larger shoal in the poeciliid fish *Girardinus falcatus*. Ethol. Ecol. Evol. 19 (2), 145–157.

Agrillo, C., Dadda, M., Bisazza, A., 2007. Quantity discrimination in female mosquitofish. Anim. Cogn. 10, 63–70.

Agrillo, C., Dadda, M., Serena, G., Bisazza, A., 2008. Do fish count? Spontaneous discrimination of quantity in female mosquitofish. Anim. Cogn. 11 (3), 495–503, http://dx.doi.org/10.1007/s10071-008-0140-9.

Agrillo, C., Dadda, M., Serena, G., Bisazza, A., 2009. Use of number by fish. PLoS One 4 (3), e4786.

Agrillo, C., Miletto Petrazzini, M.E., Bisazza, A., 2014. Numerical acuity of fish is improved in the presence of moving targets, but only in the subitizing range. Anim. Cogn. 17 (2), 307–316.

Agrillo, C., Miletto Petrazzini, M.E., Piffer, L., Dadda, M., Bisazza, A., 2012a. A new training procedure for studying discrimination learning in fishes. Behav. Brain Res. 230, 343–348.

Agrillo, C., Miletto Petrazzini, M.E., Tagliapietra, C., Bisazza, A., 2012b. Inter-specific differences in numerical abilities among teleost fish. Front. Psychol. 3 (483), http://dx.doi.org/10.3389/fpsyg.2012.00483.

Agrillo, C., Piffer, L., Bisazza, A., 2011. Number versus continuous quantity in numerosity judgements by fish. Cognition 119, 281–287.

Agrillo, C., Piffer, L., Bisazza, A., Butterworth, B., 2012c. Evidence for two numerical systems that are similar in humans and guppies. PLoS One 7 (2), e31923.

Agrillo, C., Piffer, L., Bisazza, A., Butterworth, B., 2015. Ratio dependence in small number discrimination is affected by the experimental procedure. Front. Psychol. 6 (1649), http://dx.doi.org/10.3389/fpsyg.2015.01649.

Arsalidou, M., Taylor, M.J., 2011. Is 2 + 2 = 4?: Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage* 54, 2382–2393.

Aronson, L.R., 1956. Further studies on orientation and jumping behavior in the goby fish, *Bathygobius soporator*. Anat. Rec. 125, 606.

Beran, M.J., 2007. Rhesus monkeys (*Macaca mulatta*) enumerate large and small sequentially presented sets of items using analog numerical representations. J. Exp. Psychol.: Anim. Learn. Cogn. 33 (1), 42–54.

Beran, M.J.1, 2008. Monkeys (*Macaca mulatta* and *Cebus paella*) track, enumerate, and compare multiple sets of moving items. J. Exp. Psychol.: Anim. Behav. Process. 34 (1), 63–74.

Beran, M.J., Parrish, A.E., 2013. Visual nesting of stimuli affects rhesus monkeys' (*Macaca mulatta*) quantity judgements in a bisection task. Atten. Percept. Psychophys. 75, 1243–1251.

Bisazza, A., Agrillo, C., Lucon-Xiccato, T., 2014. Extensive training extends numerical abilities of guppies. Anim. Cogn. 7 (6), 1413–1419.

Bisazza, A., Piffer, L., Serena, G., Agrillo, C., 2010. Ontogeny of numerical abilities in fish. PLoS One 5 (11), e15516.

Bogale, B.A., Aoyama, M., Sugita, S., 2014. Spontaneous discrimination of food quantities in the jungle crow, *Corvus macrorhynchos*. Anim. Behav. 94, 73–78.

Bonanni, R., Natoli, E., Cafazzo, S., Valsecchi, P., 2011. Free-ranging dogs assess the quantity of opponents in intergroup conflicts. Anim. Cogn. 14, 103–115.

Bradner, J., McRobert, S.P., 2001. The effect of shoal size on patterns of body colour segregation in mollies. J. Fish Biol. 59 (4), 960–967.

Brown, C., 2015. Fish intelligence, sentience, and ethics. Anim. Cogn. 18, 1–17, http://dx.doi.org/10.1007/s10071-014-0761-0.

Brown, C., Laland, K., Krause, J., 2011. Fish Cognition and Behavior. Wiley-Blackwell, Oxford, UK.

Boysen, S.T., Berntson, G.G., 1989. Numerical competence in a chimpanzee (*Pan troglodytes*). J. Comp. Psychol. 103 (1), 23–31, http://dx.doi.org/10.1037/0735-7036.103.1.23.

Cantlon, J.F., Brannon, E.M., 2006. Shared system for ordering small and large numbers in a monkey. Psychol. Sci. 17, 401–406.

Cantlon, J.F., Brannon, E.M., 2007. How much does number matter to a monkey? J. Exp. Psychol. Anim. Behav. Process. 33 (1), 32–41.

Cantlon, J.F., Piantadosi, S.T., Ferrigno, S., Hughes, K.D., Barnard, A.M., 2015. The origins of counting algorithms. Psychol. Sci. 26 (6), 853–865, http://dx.doi.org/10.1177/0956797615572907.

Cordes, S., Brannon, E.M., 2008. Quantitative competencies in infancy. Dev. Sci. 11, 803–808.

Cordes, S., Gelman, R., Gallistel, C.R., Whalen, J.1, 2001. Variability signatures distinguish verbal from nonverbal counting for both small and large numbers. Psychon. Bull. Rev. 8, 698–707.

Dadda, M., Agrillo, C., Bisazza, A., Brown, C., 2015. Laterality enhances numerical skills in the guppy, *Poecilia reticulata*. Front. Behav. Neurosci. 9 (285), http://dx.doi.org/10.3389/fnbeh.2015.00285.

Dadda, M, Piffer, L., Agrillo, C., Bisazza, A., 2009. Spontaneous number representation in mosquitofish. Cogn. 112, 343–348.

Davis, H., 1984. Discrimination of the number three by a raccoon (*Procyon lotor*). Anim. Learn. Behav. 4 (12), 409–413.

DeLong, C.M., Keller, A.M., Wilcox, K.T., Keenan, S.A., (2016). Visual discrimination of geometric and complex 2D shapes in goldfish (*Carassius auratus*). Manuscript submitted for publication.

DeWind, N., Brannon, E.M., 2012. Malleability of the approximate number system: effects of feedback and training. Front. Hum. Neurosci. 6, 68 (10.339/fnhum.2012.00068).

Emmerton, J., Remmer, J.C., 2009. Local rather than global processing of visual arrays in numerosity discrimination by pigeons (*Columba liva*). Anim. Cogn. 12, 511–526.

Feigenson, L., Dehaene, S., Spelke, E.S., 2004. Core systems of number. Trends Cogn. Sci. 8, 307–314, http://dx.doi.org/10.1016/j.tics.2004.05.002.

Gellermann, L.W., 1933. Chance orders of alternating stimuli in visual discrimination experiments. J. Genet. Psychol. 42 (1), 206–208.

Gomez-Laplaza, L., Gerlai, R., 2011a. Can angelfish (*Pterophyllum scarlare*) count? Discrimination between different shoal sizes follows Weber's Law. Anim. Cogn. 14, 1–9.

Gomez-Laplaza, L.M., Gerlai, R., 2011b. Spontaneous discrimination of small quantities: shoaling preferences in angelfish (*Pterophyllum scalare*). Anim. Cogn. 14, 65–574.

Gomez-Laplaza, L.M., Gerlai, R., 2013. Quantification abilities in angelfish (*Pterophyllum scalare*): the influence of continuous variables. Anim. Cogn. 16, 373–383, http://dx.doi.org/10.1007/s10071-012-0578-7.

Hager, M.C., Helfman, G.S., 1991. Safety in numbers: shoal size choice by minnows under predatory threat. Behav. Ecol. Sociobiol. 29 (4), 271–276.

Hennig, M., 1977. Counting ability in Blennius-pavo risso (*Blenniidae, perciformes*). Zool. Anz. 199, 1–18.

Hommel, G., 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 75 (2), 383–386, http://dx.doi.org/10.1093/biomet/75.2.383.

Hosmer, D.W., Lemeshow, S., 1980. Goodness of fit tests for the multiple logistic regression model. Commun. Stat.—Theory Methods 9 (10), 1043–1069.

Hunt, S., Low, J., Burns, K.C., 2008. Adaptive numerical competency in a food-hoarding songbird. Proc. R. Soc. Lond. B 267, 2373–2379.

Irie, N., Hasegawa, T., 2012. Summation by asian elephants (*Elephas maximus*). Behav. Sci. 2, 50–56, http://dx.doi.org/10.3390/bs2020050.

Irie-Sugimoto, N., Kobayashi, T., Sato, T., Hasegawa, T., 2009. Relative quantity judgment by asian elephants (*Elephas maximus*). Anim. Cogn. 12, 193–199, http://dx.doi.org/10.1007/s1007-008-0185-9.

Izard, V., Sann, C., Spelke, E.S., Streri, A., 2009. Newborn infants perceive abstract numbers. Proc. Nat. Acad. Sci. U. S. A. 106, 10382–10385, http://dx.doi.org/10.1073/pnas.081214216.

Jaakkola, K., Fellner, W., Erb, L., Rodriguez, M., Guarino, E., 2005. Understanding of the concept of numerically less by bottlenose dolphins (*Tursiops truncatus*). J. Comp. Psychol. 119, 296–303, http://dx.doi.org/10.1037/0736-7036.119.3.296.

Keller, A.M., DeLong, C.M., 2014. Numerical discrimination in goldfish (*Carassius auratus*). In: Poster Presented at the 21st Annual International Conference on Comparative Cognition, Melbourne, FL.

Kilian, A., von Fersen, L., Gunturkun, O., 2005. Left hemispheric advantage for numerical abilities in the bottlenose dolphin. Behav. Process. 68, 179–184, http://dx.doi.org/10.1016/j.beproc.2004.11.003.

Kilian, A., Yama, S., von Fersen, L., Gunturkun, O., 2003. A bottlenose dolphin discriminates visual stimui differing in numerosity. Learn. Behav. 31 (2), 133–142.

Krusche, P., Uller, C., Dicke, U., 2010. Quantity discrimination in salamanders. J. Exp. Biol. 213, 1822–1828.

Lopez, J.C., Broglio, C., Rodriguez, F., Thinus-Blanc, C., Salas, C., 1999. Multiple spatial learning strategies in goldfish (*Carassius auratus*). Anim. Cogn. 2, 109–120.

Lucon-Xiccato, T., Bisazza, A., 2016. Male and female guppies differ in speed but not in accuracy in visual discrimination learning. Anim. Cogn. 19, 733–744.

Lucon-Xiccato, T., Dadda, M., Bisazza, A., 2016. Sex differences in discrimination of shoal size in the guppy (*Poecilia reticulata*). Ethology 122 (6), 481–491.

Lucon-Xiccato, T., Petrazzini, M.E.M., Agrillo, C., Bisazza, A., 2015. Guppies discriminate between two quantities of food items but prioritize item size over total amount. Anim. Behav. 107, 183–191.

Mandler, G., Shebo, B.J., 1982. Subitizing: an analysis of its component processes. J. Exp. Psychol. Gen. 111 (1), 1–22, http://dx.doi.org/10.1037/0096-3445.111.1.1.

McAroe, C.L., Craig, C.M., Holland, R.A., 2016. Place vs. response learning in fish: a comparison between species. Anim. Cogn. 19, 153–161, http://dx.doi.org/10.1007/s10071-015-0922-9.

McComb, K., Packer, C., Pusey, A., 1994. Roaring and numerical assessment in contests between groups of female lions, *Panthera leo*. Anim. Behav. 47 (2), 379–387.

Mehlis, M., Thunkun, T., Bakker, T.C.M., Frommen, J.G., 2015. Quantification acuity in spontaneous shoaling decisions of three-spined sticklebacks. Anim. Cogn. 18, 1125–1131, http://dx.doi.org/10.1007/s10071-015-0884-y.

Mesquita, F.O., Borcato, F.L., Huntingford, F.A.1, 2015. Cue-based and algorithmic learning in common carp: a possible link to stress coping style. Behav. Process. 115, 25–29.

Miletto Petrazzini, M.E., Agrillo, C., 2016. Turning to the larger shoal: are there individual differences in small-and large-quantity discrimination of guppies? Ethol. Ecol. Evol. 28 (2), 211–220.

Miletto Petrazzini, M.E., Agrillo, C., Piffer, L., Bisazza, A., 2014. Ontogeny of the capacity to compare discrete quantities in fish. Dev. Psychobiol. 56 (3), 2014, 529–536.

Miletto Petrazzini, M.E., Agrillo, C., Piffer, L., Dadda, M., Bisazza, A., 2012. Development and application of a new method to investigate cognition in newborn guppies. Behav. Brain Res. 233, 443–449.

Mora-Ferrer, C., Neumeyer, C., 2009. Neuropharmacology of vision in goldfish: a review. Vis. Res. 49 (9), 960–969.

Odic, D., Hock, H., Halberda, J., 2014. Hysteresis affects approximate number discrimination in young children. J. Exp. Psychol.: Gen. 143, 255–265, http://dx.doi.org/10.1037/a0030825.

Pepperberg, I.M., 1994. Numerical competence in an African gray parrot (*Psittacus erithacus*). J. Comp. Psychol. 108 (1), 36–44, http://dx.doi.org/10.1037/0735-7036.108.1.36.

Pepperberg, I., 2006. Gray parrot numerical competence: a review. Anim. Cogn. 9, 377–391, http://dx.doi.org/10.1007/s10071-006-0034-7.

Perdue, B.M., Talbot, C.F., Stone, A.M., Beran, M.J., 2012. Putting the elephant back in the heard: elephant relative quantity judgements match those of other species. Anim. Cogn. 15, 955–961.

Piazza, M., Mechelli, A., Butterworth, B., Price, C.J., 2002. Are subitizing and counting implemented as separate or functionally overlapping processes? NeuroImage 15, 435–446.

Piffer, L., Agrillo, C., Hyde, C.D., 2012. Small and large number discrimination in guppies. Anim. Cogn. 15, 215–221.

Piffer, L., Miletto Petrazzini, M.E., Agrillo, C., 2013. Large number discrimination in newborn fish. PLoS One 8 (4), e62466.

Pisa, E.P., Agrillo, C., 2009. Quantity discrimination in felines: a preliminary investigation of the domestic cat (*Felis silvestris catus*). J. Ethol. 27 (2), 289–293.

R Core Team, 2016. R: A Language and Environment for Statistical Computing (3.3.0) [programming Language]. R Foundation for Statistical Computing, Vienna, Austria (Retrieved from https://www.R-project.org/).

Reznikova, Z., Ryabko, B., 2011. Numerical competence in animals, with an insight from ants. Behaviour 148 (4), 405–434.

Rodriguez, R.L., Briceno, R.D., Briceno-Aguilar, E., Hobel, G., 2015. *Nephila clavipes* spiders (araena: Nephilidae) keep track of captured prey counts: testing for a sense of numerosity in an orb-weaver. Anim. Cogn. 18, 307–314.

Rugani, R., Regolin, L., Vallortigara, G., 2008. Discrimination of small numerosities in young chicks. J. Exp. Psychol. Anim. Behav. Process 34 (3), 388–399.

Schwarz, G.E., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464, http://dx.doi.org/10.1214/aos/1176344136.

Scheffé, H., 1959. The Analysis of Variance. Wiley, New York.

Siebeck, U.E., Litherland, L., Wallis, G.M., 2009. Shape learning and discrimination in reef fish. J. Exp. Biol. 212, 2113–2119 (10.124/jeb.028936).

Srinivasan, M.V., 2010. Honeybees as a model for vision, perception, and cognition. Ann. Rev. Entomol. 55, 267–284, http://dx.doi.org/10.1146/annurev.ento.010908.164537.

Stancher, G., Rugani, R., Regolin, L., Vallortigara, G., 2015. Numerical discrimination by frogs (*Bombina orientalis*). Anim. Cogn. 18, 219–229.

Uller, C., Jaeger, R., Guidry, G., Martin, C., 2003. Salamanders (*Plethodon cinereus*) go for more: rudiments of number in a species of basal vertebrate. Anim. Cogn. 6, 105–112.

Tomonaga, M., 2008. Relative numerosity discrimination by chimpanzees (Pan troglodytes): Evidence for approximate numerical representations. Anim. Cogn. 11, 43–57.

Trick, L.M., Pylyshyn, Z.W., 1994. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psychol. Rev. 101, 80–102, http://dx.doi.org/10.1037/0033-295X.101.1.80.

Venables, W.N., Ripley, B.D., 2002. MASS: Support Functions and Datasets for Venables and Ripley's MASS (7.3-45) [R Package] (Retrieved from https://CRAN.R-project.org/package=MASS).

Vonk, J., 2014. Quantity matching by an orangutan. Anim. Cogn. 17, 297–306, http://dx.doi.org/10.1007/s10071-013-0662-7.

Wyzisk, K., Neumeyer, C., 2007. Perception of illusory surfaces and contours in goldfish. Vis. Neurosci. 25, 291–298.

Xu, F., Spelke, E.S., 2000. Large number discrimination in 6-month-old infants. Cognition 74, B1–B11, http://dx.doi.org/10.1016/S0010-0277(99)00066-9.

Yaman, S., Kilian, A., von Fersen, L., Gunturkun, O., 2012. Evidence for a numerosity category that is based on abstract qualities of few vs. many in the bottlenose dolphin (Tursiops truncatus). Front. Psychol. 3 (00473), http://dx.doi.org/10.3389/fpsyg.2012.00473.