Anim Cogn (2011) 14:695–706
DOI 10.1007/s10071-011-0404-7

ORIGINAL PAPER

Relative quantity judgments in South American sea lions (*Otaria flavescens*)

José Z. Abramson · Victoria Hernández-Lloreda · Josep Call · Fernando Colmenares

Received: 14 October 2010 / Revised: 7 March 2011 / Accepted: 11 April 2011 / Published online: 28 April 2011

Abstract There is accumulating evidence that a variety of species possess quantitative abilities although their cognitive substrate is still unclear. This study is the first to investigate whether sea lions (*Otaria flavescens*), in the absence of training, are able to assess and select the larger of two sets of quantities. In Experiment 1, the two sets of quantities were presented simultaneously as whole sets, that is, the subjects could compare them directly. In Experiment 2, the two sets of quantities were presented item-by-item, and the totality of items was never visually available at the time of choice. For each type of presentation, we analysed the effect of the ratio between quantities, the difference between quantities and the total number of items presented. The results showed that (1) sea lions can make relative quantity judgments successfully and (2) there is a predominant influence of the ratio between quantities on the subjects' performance. The latter supports the idea that an analogue representational mechanism is responsible for sea lions' relative quantities judgments. These findings are consistent with previous reports of relative quantities judgments in other species such as monkeys and apes and suggest that sea lions might share a similar mechanism to compare and represent quantities.

Keywords Numerical cognition · Relative quantity judgment · Sea lion · Comparative cognition · Accumulator model

The experiments comply with the current laws of Spain.

J. Z. Abramson (✉) · F. Colmenares
Dpto. Psicobiología, Universidad Complutense de Madrid, Madrid, Spain
e-mail: zabramson@psi.ucm.es

V. Hernández-Lloreda (✉)
Dpto. Metodología de las Ciencias del Comportamiento, Universidad Complutense de Madrid, Madrid, Spain
e-mail: vhlloreda@psi.ucm.es

J. Call
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Introduction

Many organisms face problems in their ecological and social niches in which the ability to correctly estimate quantities can be fitness enhancing. Thus, benefits associated with, for example, more efficient strategies for foraging, hunting, mating, competing and cooperating with conspecifics (Hauser 2005; Tomasello and Call 1997) may be maximised if individuals are endowed with cognitive skills to assess relative and absolute quantities. Some of these functions have already been documented or hinted at, for example, when individuals make flexible behavioural decisions in response to the number of potential prey in differently sized groups (fish: Botham et al. 2005), the number of partners in a hunting party (chimpanzees: Watts and Mitani 2002), the number of mating partners or rivals in different territories or groups (baboons: Kitchen et al. 2004) or the number of rivals in intergroup encounters (chimpanzees: Wilson et al. 2007 and lions: McComb et al. 1994).

Uncovering the cognitive mechanism underlying quantity representation has become an important challenge for those interested in understanding the ontogeny and evolution of cognition (Brannon and Terrace 2002; Carey 2001; Hauser and Spelke 2004; Uller 2008; Xu and Carey 1996) Although perceptual appraisal mechanisms such as subitizing (Davis and Perusse 1988) or prototype matching (Thomas 1992) could account for some of the early findings, they cannot accommodate the discrimination of pairs of quantities that (1) fall outside of the subitizing range (Dooley and Gill 1977) and (2) are presented sequentially and therefore are not visually available at the same time (Beran 2001; Call 2000; Hanus and Call 2007).

Two different mechanisms have been proposed to replace the purely perceptual mechanisms: the object file model (Kahneman and Treisman 1984; Kahneman et al. 1992; Simon et al. 1995; Uller et al. 1999) and the accumulator model (Gallistel and Gelman 2000; Meck and Church 1983). According to the object file model, numerical capacity works on mental (symbolic) representations of a set of visual objects, i.e., the object files. Each object file is a mental token that represents each element of a set of elements, yielding exactly as many files (mental tokens) as objects filed in short-term memory (Uller et al. 1999; Wiese 2003). In contrast, according to the accumulator model, animals cannot discriminate absolute numbers or label each separate object. However, they can recognise quantities by means of an accumulated analogue representation, the accumulation of continuous quantities in proportion to the number of quantified elements (Meck and Church 1983). That is, discrete quantities can be represented as mental magnitudes that could be seen as an analogue of the perceived discrete quantities.

Although the object file model does provide an accurate representation of quantities, the underlying mechanism places a serious constraint on the set size limit. Indeed, since the mechanism operating requires memory space which is limited, quantities greater than 3 or 4 are hard to process simultaneously (Uller et al. 1999). In contrast, the analogue representation (magnitude) of the accumulator model can deal with larger sets, since it has no a priori limit (Dehaene 2001; Gallistel and Gelman 2000), but is rather noisy (Gallistel and Gelman 2000), as the accuracy of discrimination decreases with increasing quantities (and the absolute difference between quantities is kept constant). Understanding the diversity (and commonality) of cognitive mechanisms supporting the various numerical skills reported in a variety of animal species, including humans, as well as the ecological and evolutionary factors that may have driven them, is a challenge that needs to be addressed with a broad comparative approach (Shettleworth 2009). The only aquatic mammal in which quantitative cognition skills have been investigated is the bottlenose dolphin (Jaakkola et al. 2005; Kilian et al. 2003, 2005). In contrast to dolphins, sea lions are shorter lived, have a less complex social organisation and are less encephalized (Reichmuth Kastak and Schusterman 2002). Nevertheless, it is known that California sea lions, the pinniped on whom most of the cognitive research has been done so far, are relatively large-brained social mammals that learn rather quickly to perform complex behaviours in captivity (Pack et al. 1991). They have shown complex cognitive abilities in artificial sign language comprehension, short- and long-term memory, discrimination learning, associative learning, concept formation and equivalence classification (see Schusterman et al. 2002, for a review). Moreover, sea lions show special skills to trick and steal fishermen's catch (Schusterman et al. 2002).

South American sea lions (*Otaria flavescens*), the species of the present study, often forage for schooling prey in groups (Sepulveda et al. 2007). Estimating the appropriate group size needed for a coordinated successful hunt may have some adaptive value. The sea lion's mating system is another natural context in which a selective pressure might have favoured the evolution of cognitive skills devoted to assessing numerousness. Thus, during the breeding season, males establish territories that comprise on average 5–20 females (Acevedo et al. 2003). During this period, group raids happen, mostly caused by gangs of subadult males attempting to mate with the females (Capozzo 2002). Although the resident males are unable to fight off all the raiders and keep all the females in their territorial boundaries, they must consider their harem size to monopolise and protect the females from other intruder males' attempts to copulate with them.

The present study investigated numerical cognition in a pinniped, the South American sea lion (*Otaria flavescens*). In a previous study, Genty and Roeder (2006) showed that three out of four California sea lions could master a reverse-reward task using different quantities of food. Nevertheless, that study focused on self-control, not on numerical cognition. Moreover, the experimental set-up required subjecting the animals to long training sessions and the presentation of limited sets of quantities comparisons.

We wanted to (1) find out whether sea lions are able to select the larger of two sets of quantities in absence of training, (2) determine whether such discrimination can be done based on mental representations of quantity, rather than rely strictly on perceptual cues and (3) explore what representational system could underpin these capacities. In Experiment 1, the two sets of quantities were presented simultaneously as whole sets, that is, the subjects could compare them directly (see, for example, Call 2000; Hanus and Call 2007; Irie-Sugimoto et al. 2009). In Experiment 2, the two sets of quantities were presented item-by-item (Beran 2001; see also, Hanus and Call 2007; Irie-Sugimoto et al. 2009), and the totality of items was never visually available at the time of choice since subjects only witnessed the action of dropping a different number of items into one of two opaque boxes. Thus, to succeed, subjects were required to store and compare the quantities mentally. We analysed the effect of the ratio between pairs of quantities to contrast the accumulator and the object file model. An analogue system like the accumulator model predicts an influence of the ratio between quantities on the subjects' performance, whereas the object file model predicts a cut-off point in discrimination ability when quantities exceed the number of four.

Materials and methods

Subjects

We tested 4 South American sea lions (*Otaria flavescens*): Ana (A) and Laura (L), two 8-year-old females housed at L'Oceanografic of Valencia, in Spain; Erica (E), a 5-year-old female and Simon (S), a 3-year-old male, both housed at Madrid Zoo Aquarium, in Spain. Experiments were conducted in February 2009 at L'Oceanografic of Valencia and in June and July 2009 at Madrid Zoo Aquarium. Experimental sessions were conducted typically between 11:00 a.m. and 2:00 p.m. The subjects were fed approximately 4–5 kg of freshly thawed capelin each day, one half of which was typically consumed during experimental sessions. We only used herring for some pretest trials (i.e. 1/0) because herring is bigger and has higher nutritional value and, therefore, its use would have limited the number of trials to be run per day. The four subjects were trained for a variety of examination and exercise behaviours, with standard operant conditioning procedures and fish reinforcement. L and A were captured from Isla de Lobos, in Uruguay, when they were 1 year old. E and S were both mother-reared in captivity. All the subjects lived in social groups of various sizes, with access to indoor and outdoor areas. Prior to this experiment, subjects had never participated in cognitive studies. Subjects were individually tested in their outdoor cages and were neither food-, nor water-deprived. If they refused to participate in a given session, we cancelled that session and rescheduled for another time.

Apparatus

Experiment 1: Whole set presentation

Two identical white window-fronted boxes (35 × 25 × 7 cm) were used to simultaneously present sets of pieces of small–medium size fish. The boxes contained six hooks where the fish were hung. The experiments were run double blind. Each stimulus box was covered by two lids, a transparent one, that let the subject to see the rewards but prevented her/him from reaching them and an opaque one, that allowed the trainer to place the boxes in front of the subject while she/he himself was blind to its content. The number of pieces of fish inside each box varied between 0 and 6. Boxes were presented next to each other on a wooden platform (80 × 37 cm) with a track (rail) where the boxes were placed. During the experiments, placement of the fish in the boxes was done by the experimenter, seated behind the trainer, out of view both of the subject and the trainer.

Experiment 2: Item-by-item presentation

The sea lions housed at L'Oceanografic were presented with the same food quantities and procedure as in Experiment 1, except that we replaced the pair of boxes for a pair of identical square-shaped opaque buckets (35 × 25 × 40 cm). The sea lions housed at the Madrid Zoo were presented with the same boxes as in Experiment 1 but, instead of being vertically placed in the track, they were placed right side up on the platform, with their opaque lids raised as occluders that prevented the subject from seeing the content.

Procedure

All subjects received a habituation period with trainers before the study and were therefore already familiar with the apparatus.

Experiment 1: Whole set presentation

Subjects were presented with one to three daily sessions comprising between 10 and 12 two-alternative forced-choice trials. For this first experiment, testing was non-corrective and subjects received whichever quantity they happened to choose.

Prior to the test sessions, subjects received pretest trials in which the subject was shown two boxes; one box was baited and the other was empty. They were trained to come and touch or approach the baited box to receive fish. When they reached 80% per cent of correct choices, we considered that they were ready to be tested. In addition, to rule out lack of motivation prior to testing, we required the subject to undergo 1-0 pretest trials prior to the test and to choose the baited box for three consecutive trials. Once the subject met this criterion, we administered the test trials in which both boxes were baited with different amount of fish.

The experimenter filled each box with a certain amount of fish out of view from both the subject and the trainer, covered each box with the opaque lid and handed them over to the trainer who placed the two boxes into the rail of the platform. The boxes were next to each other at a distance of 5 cm from the centre of the platform to the edge of each box and approximately 50 cm away from the subject. Next, the trainer uncovered the boxes showing their content to the subject. The two boxes were placed in front of the trainer, in a position from which she/he could not see the content, thus avoiding any kind of possible inadvertent cuing by the trainer. Once the subject made his/her choice, the trainer covered the unselected box and gave the content of the chosen box to the subject. The subjects received the amount of fish placed in the box touched first.

Usually, twelve comparisons were presented to each subject in each session: (1 vs. 2; 1 vs. 3; 1 vs. 4; 1 vs. 5; 2 vs. 3; 2 vs. 4; 2 vs. 5; 2 vs. 6; 3 vs. 4; 3 vs. 5; 3 vs. 6 and 4 vs. 6). The interval between trials was approximately 10 s. The subject's response was considered correct when she/he chose the box with the larger quantity. At the end of each trial, the two boxes were covered and the next trial was prepared. The order of the quantities was randomly determined and counterbalanced on right and left sides with the larger amount never appearing more than twice in a row on the same side. Trials were repeated when the subject did not make a clear choice (she/he did not touch or approach any box or did not go from the waiting place to the testing arena). If subjects lost interest during presentation (when trainers indicated that a subject was tired and/or unwilling to participate), the trial was cancelled and restarted on the next day. Six trials were the minimum number of presentations in a session to make it useable (half of the complete session). All the sessions reached the criteria.

Each subject received 6 presentations of each comparison. A session was considered complete when the 12 comparisons were run or when trainers indicated that a subject was tired and/or unwilling to participate. So, depending on the subject's attention span, each subject received between 6 and 16 trials per session, totalling 4–6 testing days per subject altogether. Owing to differences in the facilities between the two aquariums, two slightly different procedures were followed.

Experimental set-up at L'Oceanografic

The subjects were tested in a separate section of the facility, an outdoor enclosure that contained a saltwater pool and an adjacent outdoor area, separated from the former by a fence, with a deck where the experimental apparatus was placed (Fig. 1a). Subjects were positioned across from the trainer separated by a steel mesh through which she/he could indicate his/her choices by approaching or touching one of the two boxes. To start each trial, the subject approached the centre of the two boxes and maintained the attention to the trainer. Once she/he was in place, the trainer removed both lids simultaneously, so that the subject could see the contents of each box (but they were out of the trainer's view). As soon as the subject had watched the contents for 2–5 s, both boxes were separated, sliding them across the rail and moved within the subject's reach so that she could make his/her choice. Subjects received two pretest sessions on different consecutive days; L received ten 1-0 counterbalanced trials and A received eighteen 1-0 counterbalanced trials.

Experimental set-up at Madrid Zoo

The subjects were tested in a separate outdoor enclosure of the facility (Fig. 1b). In this case, no barrier or steel mesh separated the subject from the boxes at the time of choice. For this presentation, we required two trainers, one would present the different amounts of fish and would give the content of the chosen box and the other would maintain the subject in a waiting point from which it was impossible for the subject to see the placement of the different quantities of fish in the boxes. The subject waited at the waiting point (playing and doing exercises) until she/he was signalled by the first trainer (who already had placed the boxes into the rail and removed both lids) to go to the testing area to make a choice. She/he responded by moving from the waiting area to the testing location where she/he would touch one of the two boxes with his/her nose. As the subject approached the boxes, the second trainer separated the two boxes, sliding them across the rail so that she/he could make his/her choice by touching one of the two boxes with her nose (fish was always out of the second trainer's view). E received 1 day pretest session with ten 1-0 counterbalanced

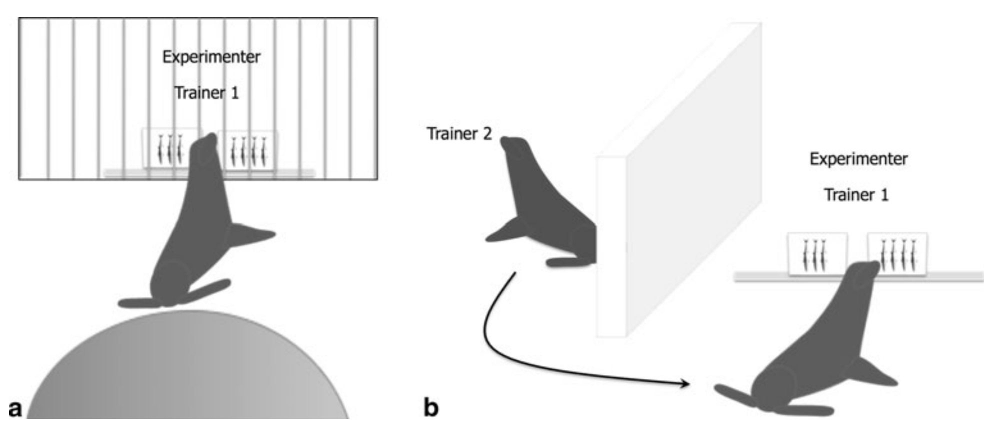

Fig. 1 Experimental set-ups: **a** at L'Oceanographic, Valencia; **b** at Madrid Zoo

trials, and S received three pretest sessions on different consecutive days, with a total of twenty-seven 1-0 counterbalanced trials.

Experiment 2: Item-by-item presentation

In Experiment 2, we investigated to what extent sea lions assess relative quantities on a more difficult task, one that, to be solved, subjects are required to mentally represent and store two quantities that were not visually accessible at the time of choice. We used the same basic procedure as in Experiment 1, except for the presentation format of the pieces of fish. In Experiment 2, instead of showing the subject the pieces of fish inside the boxes, we dropped them one-by-one into the buckets. Thus, unlike Experiment 1, subjects were unable to see the total number of pieces of fish placed inside the box.

The trainer placed both boxes/buckets on the platform in an upright position, approximately 60 cm apart in front and 70 cm away from the subject. The trainer first took out pieces of fish from a cube and dropped them one at a time into each bucket in full view of the subject, starting with the left bucket and then the right one. In doing so, subjects could only see one item at a time falling into the boxes and/or listening to the sounds made when the food hit the bottom of the bucket. The buckets were high enough so that the subject could not see their contents. Although in this case the trainer was not blind to the content, the trainer wore sunglasses and, after the presentation, looked down to avoid cuing the subject.

Once the last fish was dropped into the bucket, the trainer moved the platform (and the buckets) within the subject's reach so that she/he could make his/her choice. The trainer took away the unselected box and gave to the subject the amount of fish placed in the chosen box.

Two slightly different procedures were followed in the two aquariums.

Experimental set-up at L'Oceanografic

As in Experiment 1, prior to the test phase and prior to each testing session, we required the subject to undergo pretest trials. In these pretest trials, the subject was presented with the two buckets. The trainer first took out 1 piece of fish from a cube and dropped it into one of the buckets in full view of the subject. The other bucket remained unbaited. Subjects received two pretest sessions on consecutive days; L received fifty 1-0 counterbalanced trials and A received sixty-two 1-0 counterbalanced trials. Prior to each testing session, we required the subject to choose the baited bucket for three consecutive trials. Once the subject met the criterion, we began the test trials. Each subject received 5 presentations of each comparison.

Experimental set-up at the Madrid Zoo

Our experience with the sea lions housed at L'Oceanografic made us think that the subject might have a problem with understanding the choice component of the experimental task, partly because animals participating in public shows are routinely trained via operant conditioning not to choose but to display acrobatics and other bizarre behaviours. These individuals may confront problem-solving tasks in the same way as they confront their training for shows. Rather than trying to solve the problem spontaneously by choosing freely between options, they may wait to see what the trainer expects them to do. Therefore, in this more demanding task, we decided to reward subjects only if they chose the bucket with the larger quantity. Additionally, we changed the quantities used during the pretest because we wanted to avoid the possibility that in this harder task subjects could have learned a rule based on avoiding the empty bucket. Recall that in the pretest of Experiment 1, we used the quantities one versus zero pairing. Therefore, in the pretest of Experiment 2, we presented one piece of fish in one bucket and six pieces in the other bucket, two of the easiest quantities to discriminate in Experiment 1. E received forty-six 1-6 counterbalanced trials during three pretest sessions on different consecutive days. Unfortunately, Simón was taken to another zoo facility near a construction site and socially housed with a larger group of new companions. It became clear that the new situation prevented Simon from paying adequate attention to the task.

Data scoring and analysis

All trials were videotaped from a distance of 10 metres to the subject with a digital full HD 1080 Sony HDR-XR520 camera. As sample size was small ($n = 4$), data were analysed, except for the mediation analysis—see below—at an individual level. To count as a correct response, subjects had to choose the set with the larger quantity. To address the issue of whether subjects succeeded in making relative quantity judgments, that is, if subjects performed above chance levels ($P = 0.5$), a dependent dichotomous variable, scored 1 for a correct response (larger quantity selected) and 0 for an incorrect response (smaller quantity selected), was used. Binomial tests for each subject were performed.

To address the issue of whether performance varied with (1) disparity: *difference* (larger quantity minus smaller quantity) and *ratio* (smaller quantity/larger quantity) and (2) total quantity (smaller quantity + larger quantity), Pearson correlation coefficients were calculated for each subject between these three variables and the proportion of correct responses (proportion of trials in which subjects selected the larger quantity of the pair). In order to test whether the data matched the prediction from the object file model, we divided the trials into two categories: below and above the limit predicted by the object file model (i.e. 4 items in a set; see, e.g. Hauser et al. 2000; Uller et al. 1999).

In order to rule out the possibility that sea lions have an object file limit different from 4 because of differences in memory skills, we explored the subjects' performance for each comparison by looking for some indication of a set size limitation (such as an abrupt change in performance when tested for comparisons over or below a particular quantity).

The point-biserial correlation coefficient was calculated for each subject between the proportion of correct trials and this categorical variable. Following correlation analyses, we analysed potential mediation effects among the variables that showed a significant correlation with performance. To determine whether mediation occurs (see Baron and Kenny 1986; Kenny et al. 1998), four criteria must be met in turn. First, a significant correlation between the predictor and the outcome must be found. Second, a significant correlation between the predictor and the mediator must be found. Here, we used the proportion of trials in which the subject chose the larger quantity of the pair as our dependent variable. Third, the mediator must be found to affect the outcome when the predictor is controlled for. Finally, it must be determined whether complete or partial mediation has occurred; complete mediation is indicated by the effect of the predictor on the outcome being completely removed when the mediator is controlled for. As for the third and fourth criteria, a regression analysis was run, the outcome being the dependent variable (here, quantities discrimination performance) and with the mediator and predictor entered simultaneously as independent variables.

If the first three criteria are satisfied but the fourth is not, partial mediation is indicated. If all the four criteria are satisfied, a total mediation effect is supported. In addition, Sobel tests (Sobel 1982) were conducted to assess the significance of mediation effects. A second observer scored 40% of the sessions to assess inter-observer reliability. Inter-observer reliability was excellent (Cohen's kappa = 0.99, $P < 0.001$).

Results

Experiment 1

Table 1 presents (1) the proportion of correct responses for each subject, (2) the Pearson correlation coefficients (r) with the three predictors investigated: the *ratio* (smaller quantity/larger quantity), the *difference* (larger quantity minus smaller quantity) and the *total quantity* (smaller quantity + larger quantity) and (3) the point-biserial correlation coefficient (r_{pb}) with the object file dichotomous variable, i.e., quantities below and above the object file limit. All subjects chose significantly more often the box with the larger quantity (binomial tests: $P < 0.05$ in all cases). Pearson correlation coefficients revealed that in three out of four subjects, performance improved as the disparity between pairs of sets increased, making the ratio the best predictor of a subject's performance followed by the difference (Figs. 2, 3). Point-biserial correlation revealed that individual performance showed no breakdown for quantity discriminations that went beyond four items, as the object file model would predict. On the contrary, although the relationship did not reach statistical significance, the results went against what the model predicts; subjects tended to fail when they had to make quantity comparisons below the object file limit.

Table 1 Percentage of correct trials for each subject, Pearson correlation coefficients (r) with the three predictors investigated—the *ratio*, the *difference*, and the *total quantity*—and the point-biserial correlation coefficient (r_{pb}), with the object file dichotomous variable—quantities below and above the object file limit—for condition 1

Subject	% Correct (P) ($n = 72$)	Ratio r (P) (All $n = 12$)	Difference r (P) (All $n = 12$)	Total r (P) (All $n = 12$)	Object file r_{bp} (P) (All $n = 12$)
L'Oceanografic					
Ana	**65 (0.013)**	0.09 (0.775)	−0.07 (0.826)	−0.03 (0.932)	0.11 (0.734)
Laura	**68 (0.003)**	**−0.77 (0.003)** $R^2 = 0.59$	**0.74 (0.006)** $R^2 = 0.55$	−0.07 (0.828)	−0.40 (0.211)
Madrid Zoo					
Erica	**92 (<0.001)**	**−0.76 (0.004)** $R^2 = 0.58$	**0.63 (0.029)** $R^2 = 0.40$	−0.13 (0.686)	−0.26 (0.418)
Simón	**81 (<0.001)**	**−0.70 (0.011)** $R^2 = 0.49$	**0.69 (0.014)** $R^2 = 0.48$	0.07 (0.820)	−0.29 (0.385)

Bold values represent statistically significant results ($P < 0.05$). For statistically significant correlations, effect size (R^2) is reported

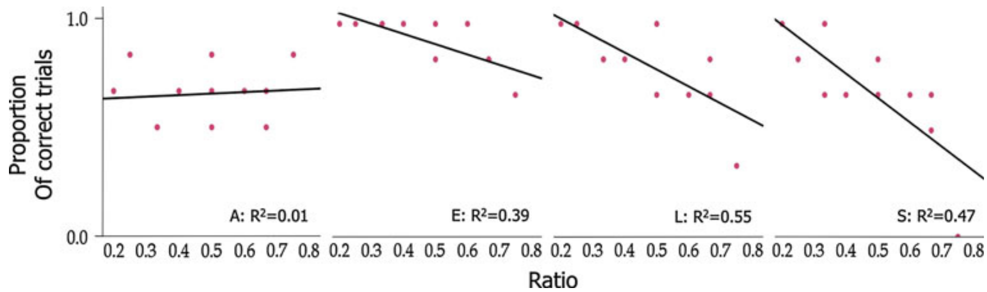

Fig. 2 Proportion of correct trials against ratio between quantities presented as pairs of whole sets (condition 1). Estimated linear regression equations and R^2 for each subject are reported

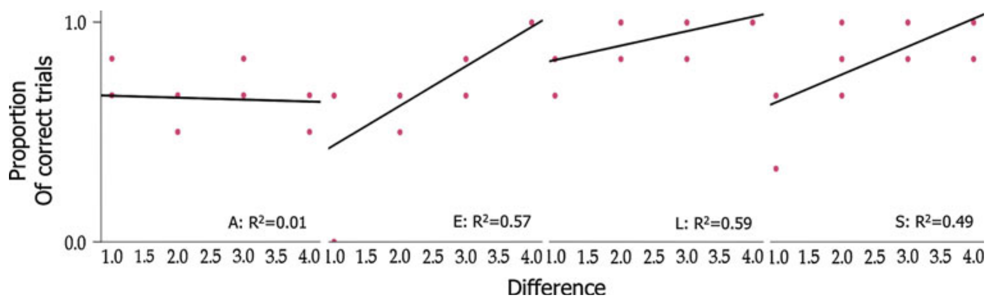

Fig. 3 Proportion of correct trials against difference between quantities presented as pairs of whole sets (condition 1). Estimated linear regression equations and R^2 for each subject are reported

Table 2 shows the subjects' performance for each comparison. Performance did not show a sudden change for any of the tested quantities, which further argues against strict set size limitations and thus against the object file model.

Successful performance was negatively correlated with ratio and positively correlated with difference. As ratio and difference were also negatively correlated with each other ($r = 0.73$, $n = 12$, $P < 0.001$), we hypothesised that the ratio between quantities would mediate the effect of the difference on performance, making the ratio the genuine factor that accounts for the pattern of subjects' errors as would be expected according to Weber's law. To determine whether mediation occurs, we evaluated the four criteria previously explained (see above). First, a significant correlation between the predictor (here, the difference between quantities) and the outcome (here, the percentage of correct choices) was found (see Table 1). Second, as we previously mentioned, a significant correlation between the predictor (here, the difference) and the mediator (here, the ratio between quantities) was found. Thus, our data fulfilled the first two criteria. The analyses were run including the three subjects that showed a significant effect of ratio and difference (L, E and S).

Table 2 Percentage of correct trials for each subject per comparison in Experiment 1

Comparison	% Correct			
	A	L	S	E
1 versus 2	67	67	67	83
1 versus 3	50	67	83	100
1 versus 4	83	83	100	100
1 versus 5	67	100	100	100
2 versus 3	67	67	67	83
2 versus 4	50	67	100	100
2 versus 5	67	67	83	100
2 versus 6	50	100	83	100
3 versus 4	83	0	33	67
3 versus 5	67	67	83	100
3 versus 6	83	83	100	83
4 versus 6	50	5	83	83

As for the third and fourth criteria, a regression analysis was run, the outcome being the dependent variable (here, quantities discrimination performance) and with the mediator (ratio) and predictor (difference) entered simultaneously as independent variables.

Consistent with criterion 3 (the mediator must be found to affect the outcome when the predictor is controlled for), ratio was found to be associated with performance when the difference was controlled for ($\beta = -0.548$, $t = -2.070$, $P = 0.04$). The results indicate total mediation; difference was no longer a significant predictor of performance when the ratio was controlled for ($\beta = 0.055$, $t = 1.270$, $P = 0.21$). In addition, there was a mediation effect, although not statistically significant following standard criteria (Sobel test = 1.818; $P = 0.06$). Nevertheless, given that the Sobel test is very conservative (MacKinnon et al. 1995), we can consider the ratio between quantities to be a total mediator of the relationship between difference and relative quantity judgments. These findings provide strong support for the hypothesis that the effect of difference on performance is due to the mediation effect of ratio. Furthermore, to reject an alternative account, that the difference between quantities might in fact mediate the negative correlation between ratio and performance, the Sobel test was performed but in this case with ratio as the predictor and the difference as the possible mediator. The mediator (here, difference) did not affect the performance when we controlled for ratio (*Sobel test* $= -0.13$; $P = 0.89$).

Experiment 2

Of the two subjects initially available at the Madrid Zoo, we could only test one because the other was later unavailable due to management decisions. Of the three tested, only one subject (E) chose the box with the larger quantity at a significantly greater frequency than the smaller one. It is important to note that the subject that succeeded was the one for which 1 vs. 6 pretest trials were used. Table 3 presents (1) the percentage of correct trials for this subject, (2) the Pearson correlation coefficients (r) with the three predictors investigated: the *ratio*, the *difference* and the *total quantity* and (3) the point-biserial correlation coefficient (r_{pb}) with the object file dichotomous variable, i.e., quantities below and above the object file limit.

Since the protocol of Experiment 2 incorporated two modifications compared to Experiment 1, i.e., a corrective procedure and a pretest in which 1 vs. 6 comparisons were used, the subject required more trials to understand the procedure. Thus, in order to rule out alternative explanations, we inspected further the results. First, in order to rule out the possibility that the subject had learned to 'avoid 1' or 'go to 6' during the pretesting procedure, we explored each individual comparison. Table 4 shows E's performance for each comparison. It can be seen that 33% of correct responses in 1 vs. 3 and 3 vs. 6 comparisons or the 100% of correct choices in 3 vs. 4 comparisons do not support the hypothesis that results can be explained by associative learning. Second, since in this experiment the procedure was corrective, we explored the level of performance over time to assess whether the subject learned to respond in this way during testing. The pattern of success over time was irregular and did not increase over time (1st session: 50%, 2nd session: 75%, 3rd session: 50%, 4th session: 82%, 5th session: 75% and 6th session: 57%) as would have been expected if a process of associative learning had driven her performance.

Discussion

Sea lions were capable of selecting the larger of two quantities both when they were presented simultaneously as pairs of visually accessible whole sets (Experiment 1) and, albeit the evidence is weaker (only one subject was above chance), when presented item-by-item and the two final pairs of sets were not visually accessible at the time of choice (Experiment 2).

The results from Experiment 1 parallel those on terrestrial mammals (e.g. Hanus and Call 2007; Irie-Sugimoto et al. 2009; Ward and Smuts 2007) and bottlenose dolphins (Jaakkola et al. 2005). All subjects chose the larger quantity above chance in the simultaneous presentation, albeit we found important individual differences in the level of performance, with results ranging from 65 to 92% of correct choices. Success in choosing the larger quantity in this experiment can be based on a psychophysical mechanism, however, as both sets were simultaneously visible to the subject when making a choice.

Table 3 Percentage of correct trials for subject E, Pearson correlation coefficients with the three predictors investigated—the *ratio*, the *difference* and the *total quantity*—and the point-biserial correlation coefficient (r_{pb}), with the object file dichotomous variable—quantities below and above the object file limit—for condition 2

Subject	% Correct (P)	Ratio r (P) (All $n = 12$)	Difference r (P) (All $n = 12$)	Total r (P) (All $n = 12$)	Object file r_{bp} (P) (All $n = 12$)
L'Oceanografic					
Ana	52 (0.897) ($n = 60$)	−0.07 (0.827)	0.18 (0.580)	0.14 (0.667)	−0.33 (0.296)
Laura	52 (0.897) ($n = 60$)	0.09 (0.778)	−0.23 (0.463)	−0.21 (0.511)	0.17 (0.604)
Madrid Zoo					
Erica	**66 (0.003)** ($n = 72$)	−0.35 (0.267)	0.46 (0.136)	0.09 (0.77)	−0.19 (0.56)

Bold values represent statistically significant results ($P < 0.05$)

Table 4 Percentage of correct trials for each subject per comparison in Experiment 2

Comparison	% Correct		
	A	L	E
1 vs. 2	80	100	67
1 vs. 3	20	20	33
1 vs. 4	60	80	83
1 vs. 5	60	40	100
2 vs. 3	40	40	17
2 vs. 4	40	20	67
2 vs. 5	40	100	100
2 vs. 6	60	20	100
3 vs. 4	40	80	100
3 vs. 5	60	20	33
3 vs. 6	60	40	33
4 vs. 6	60	60	67

In Experiment 2, one of the three subjects tested did choose the set with the larger number of items. This finding indicates that at least one of the subjects was able to keep two sets of representations in working memory and compare quantities mentally, as the rewards were no longer visible after they were placed into the buckets. This confirms results from previous studies on relative quantity discrimination in other non-human species (chimpanzees: Beran 2001, 2004; orangutans: Call 2000; Hanus and Call 2007; capuchin monkeys: Evans et al. 2009; elephants: Irie-Sugimoto et al. 2009; horses: Uller and Lewis 2009; and dogs: Ward and Smuts 2007).

The less robust results obtained in Experiment 2 can be due to a variety of factors worth examining. First, the poorer performance can reflect the fact that the task was more demanding cognitively. Indeed, in this task, the subjects had to process the sequential presentation of items, one-by-one, and had to make a choice of the set with the larger amount when no reward was visually available. Second, the diminished levels of successful performance might be due to experimental artefacts such as a lack of understanding of the task or a motivational decline, the latter perhaps also related to the greater difficulty posed by the task. Nevertheless, it is worth nothing that similar trends in performance have been reported for apes. Hanus and Call (2007) found that fewer subjects succeeded in the item-by-item presentation than the simultaneous presentation. The results shifted from all subjects being above chance (with quantities varying up to six) and 90% (quantities varying up to ten) when quantities were presented simultaneously to 26% (quantities varying up to six) and 9% (quantities varying up to ten) of the subjects in the item-by-item presentation. All individuals succeeded in the whole presentation condition, while only 30% (with quantities varying up to six) and 9% (quantities varying up to ten) passed the item-by-item one. In a recent study, Evans et al. (2009) found similar results for capuchin monkeys. In the item-by-item presentation, only half of the monkeys (three out of six) succeeded in the task and performed worse (75, 77 and 70% correct, individually) than they did in the whole set presentation (all subjects achieved 85% correct). Thus, the fact that in our three-subject study only one passed the test should not be neglected. It is also possible that the reward-based approach in Experiment 2 for the subject at Madrid Zoo could have facilitated her understanding of the task and this could explain her better performance. It is thus conceivable that with additional training and a correction procedure subjects at L'Oceanographic might have improved their performance.

A critical question in the study of the origins and adaptive value of numerical capacities in species that vary widely in their ecological conditions concerns the nature of the underlying cognitive systems; the extent to which their grasping of numerousness related problems are homologous or analogous (Uller 2008). Although it is still unclear which model/s underlie relative quantitative judgments in non-human animals (Anderson et al. 2007; Beran 2001, 2004, 2007; Brannon and Terrace 2000; Hanus and Call 2007; Hauser and Carey 1998), previous studies in various species have shown that the capacity to discriminate the larger of two quantities is limited by the disparity and the magnitude of the comparisons—perhaps with the one exception found in elephants, for whom no disparity effect has been shown (Irie-Sugimoto et al. 2009). This trend strongly suggests that the accumulator model is a better candidate as a general mechanism responsible for this capacity. Our Experiment 1 shows that the errors committed by 3 out of 4 subjects are associated with the disparity (ratio) between the two quantities presented, successful performance dropping as a function of increasing ratio between quantities (ratio accounting for from 47 to 59% individual performance). Although we also found a correlation between difference and performance, the mediation analysis provided evidence for the suggestion that the effect of difference on performance is due to the mediation effect of ratio. So, our results are consistent with predictions from the accumulator model, as success declined as the ratio between quantities increased. We found no effect of total amount of pieces. Even more, individual performance did not show any breakdown for quantity discriminations that went beyond four items, as the object file model would predict. This suggests that small and large quantities are processed in much the same way and probably the object file model cannot account for the results obtained.

As for the one subject that succeeded in relative quantity judgment of the successively presented quantities in Experiment 2, although not statistically significant, we also found a similar tendency, a disparity effect in the level of her performance. Clearly, more data on successful performance in an item-by-item presentation would be needed to confirm this finding.

To solve the problems presented in experiments 1 and 2, the subjects demonstrated the capacity to deal with spatial (Experiment 1) and temporal (Experiment 2) discontinuity between food items when judging the relative quantity among set of items. Nevertheless, our study does not preclude judgments based on quantitative cues other than numerousness. For example, factors such as surface area, in Experiment 1, or dropping duration, in Experiment 2, are confounded with numerousness. Thus, in a strict sense, we cannot claim that subjects have the ability to number appreciation, even though the use of these quantitative cues might, on average, yield the same results as would numerosity (Anderson et al. 2007; Mix et al. 2002). Controlling for those factors would allow a more precise assessment of this potential numerical competence (e.g. Jaakkola et al. 2005). Nevertheless, in return, a more robust experimental design might lead to a loss of ecological validity and might also preclude a proper comparison with other studies that have used a similar protocol.

The aim of this study was to simulate problems that subjects may encounter in nature, such as fish schools swimming in different directions or going through a cave out of sight one-by-one. In studying issues of homology or convergence of cognitive traits, such as numerical cognition, measuring spontaneous behaviour is important. The use of 'non-natural tasks' that would require training may bias the results as they could reflect the operation of a process developed during the experiment rather than a process naturally available to the animal prior to the experiment (Uller and Lewis 2009). Besides, the use of closely matched experiments to compare numerical cognitive abilities across species provides us with a powerful tool in tracing the evolutionary roots of numerical cognition (Uller 2008).

Even though relative quantity judgments in these experiments were not singularly related to numerosity, and even though subjects could be using the timing of presentation or the area occupied by the fishes as the main cues, our results show that sea lions had the ability to encode and mentally compare sets of quantities, their performance increasing in accuracy as ratio between quantities declines. Furthermore, the accumulator model, proposed to describe relative numerousness judgments, could also describe relative quantitative judgments of other types of quantitative variables (such as surface, volume, duration.) because the model represents numerosity approximately as magnitudes (Beran 2004; Gallistel and Gelman 2000). Future studies should administer comparisons with larger quantities to further evaluate the predictions of the accumulator model.

For now, it is plausible to assume that comparisons across quantities, either continuous or discrete, large or small, may be based on similar quantitative mechanisms, which perhaps constitute the precursors of more sophisticated numerical skills found in other species, including humans. In this vein, Cantlon and Brannon (2006) provide evidence that this single nonverbal, evolutionarily primitive mechanism for representing and comparing numerical values as psychological magnitudes is shared for ordering small and large numbers in monkeys and humans. And recent neural evidence supports the idea that numerical discrimination in human infancy is also ratio dependent and follows Weber's Law, thus indicating continuity of these cognitive processes over development (Libertus et al. 2009). Numerical cognition stands at the core of unique human cognition achievements but its evolutionary origins remains still an open question (Uller 2008; Shettleworth 2010). Here, we present evidence of spontaneous propensity to select the greater of two quantities, a capacity that has been considered at the roots of symbolic counting (Carey 2001), in a species not previously studied, the South American sea lion (*Otaria flavescens*). This study thus adds to the set of relatively large-brained, social mammals examined for quantitative cognition and provides new information on the conceptual and representational abilities that can be at the roots of numerical competence in nonhuman animals.

Acknowledgments We are grateful to the directors of the L'Oceanografic aquarium of Valencia and Zoo Aquarium of Madrid for allowing us to conduct this research. Special thanks go to Miguel Bueno Brinkman and Manuel Castellote for their help and support and to José Antonio Esteban for help with video recordings and for providing very helpful ideas that contributed to the design of this study. We appreciate the work of the animal keepers from the two aquariums. From L'Oceanografic aquarium of Valencia, head coach Victor Sánchez, for sea lions training and management and the trainers team; Jerome Coello, Javier Escudero, Ana Coba, Luciana Suárez and Sandra Salido. Zoo Aquarium of Madrid head coach Carlos de las Parras for sea lions training and management and the trainers team; Arancha Sánchez, Cacha Martínez, Berta Romeral, Pablo Juárez, Pablo Roy, Antonio Martínez, Irene Ayllon and Craig Allum. In addition, we thank Luis Franco for his enormous help in the design and manufacturing of the apparatus used in this study. Thanks to Rosario Martínez Arias and Ma José Hernández-Lloreda for their comments and statistical advice. Special thanks go to Titaví for her artwork. Research reported in this study was partly supported by FPI studentship to José Francisco Zamorano Abramson, and project grants CCG08-UCM/SAL-4007 (Universidad Complutense de Madrid y Comunidad de Madrid) to M. V. Hernández-Lloreda, and UCM-BSCH GR58/08 (Universidad Complutense de Madrid y Banco Santander Central Hispano) and D/018712/08 AECID (Agencia Española de Cooperación Internacional y Desarrollo), Spain to F. Colmenares.

Conflict of interest The authors declared that they had no conflicts of interests with respect to their authorship and/or the publication of this article.

References

Acevedo J, Aguayo A, Sielfeld W (2003) Eventos reproductivos del león marino común (*Otaria flavescens*) (Shaw 1800), en el norte de Chile (Pacífico suroriental). Rev Biol Mar Oceanogr 38(2): 69–75

Anderson US, Stoinski TS, Bloomsmith MA, Maple TL (2007) Relative numerousness judgment and summation in Young, middleaged, and older adult Orangutans (*Pongo pygmaeus ablii* and *Pongo pygmaeus pygmaeus*). J Comp Psychol 121:1–11

Baron R, Kenny DA (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 61(6): 1173–1182

Beran MJ (2001) Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). J Comp Psychol 115:181–191

Beran MJ (2004) Chimpanzees (*Pan troglodytes*) respond to nonvisible sets after one-by-one addition and removal of items. J Comp Psychol 118:25–36

Beran MJ (2007) Rhesus monkeys (*Macaca mulatta*) succeed on computerized test designed to assess conservation of discrete quantity. Anim Cogn 10:37–45

Botham MS, Kerfoot CJ, Louca V, Krause J (2005) Predator choice in the field; grouping guppies, Poecilia reticulate, receive more attacks. Behav Ecol Sociobiol 59:181–184

Brannon EM, Terrace HS (2000) Representation of the numerosities 1–9 by rhesus macaques (*Macaca mulatta*). J Exp Psychol Anim Behav Process 26:31–49

Brannon EM, Terrace HS (2002) The evolution and ontogeny of ordinal numerical ability. In: Bekoff M, Allen C, Burghardt GM (eds) The cognitive animal: empirical and theoretical perspectives on animal cognition. The MIT Press, Cambridge, pp 197–204

Call J (2000) Estimating and operating on discrete quantities in orangutans (*Pongo pygmaeus*). J Comp Psychol 114:136–147

Cantlon JF, Brannon EM (2006) Shared system for ordering small and large numbers in monkeys and humans. Psychol Sci 17:401–406

Capozzo H (2002) South American sea lion (*Otaria flavescens*). In: Perrin W, Wursig B, Thewissen J (eds) Encyclopedia of marine mammals. Academic Press, New York, pp 1143–1146

Carey S (2001) Cognitive foundations of arithmetic: evolution and ontogenesis. Mind Lang 16:37–55

Davis H, Perusse R (1988) Numerical competence in animals: definitional issues, current evidence, and a new research agenda. Behav Brain Sci 11:561–615

Dehaene S (ed) (2001) The cognitive neuroscience of consciousness. MIT Press, Cambridge

Dooley GB, Gill TV (1977) Acquisition and use of mathematical by a linguistic chimpanzee. In: Rumbaugh DM (ed) Language learning by a chimpanzee: the Lana project. Academic Press, New York, pp 247–260

Evans T, Beran M, Harris E, Rice D (2009) Quantity judgments of sequentially presented food items by capuchin monkeys (*Cebus apella*). Anim Cogn 12:97–105

Gallistel CR, Gelman R (2000) Non-verbal numerical cognition: from reals and integers. Trends Cogn Sci 4:59–65

Genty E, Roeder J (2006) Self-control: why should sea lions (*Zalophus californianus*), perform better than primates? Anim Behav 72:1241–1247

Hanus D, Call J (2007) Discrete quantity judgments in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus*): the effect of presenting whole sets versus item-by-item. J Comp Psychol 121:241–249

Hauser MD (2005) Our chimpanzee mind. Nature 437:60–63

Hauser MD, Carey S (1998) Building a cognitive creature from a set of primitives: evolutionary and developmental insights. In: Cummins D, Allen C (eds) The evolution of mind. Oxford University Press, Oxford, pp 51–106

Hauser MD, Spelke E (2004) Evolutionary and developmental foundations of human knowledge: a case study of mathematics. In: Gazzaniga MS (ed) The cognitive neurosciences III. The MIT Press, Cambridge, pp 853–864

Hauser MD, Carey S, Hauser LB (2000) Spontaneous number representation in semi-free-ranging rhesus monkeys. Proc Biol Sci 267(1445):829–833

Irie-Sugimoto N, Kobayashi T, Sato T, Hasegawa T (2009) Relative quantity judgment by Asian Elephants (*Elephas maximus*). Anim Cogn 12(1):193–199

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: 286–303

Kahneman D, Treisman A (1984) Changing views of attention and automaticity. In: Parasuraman R, Davies D (eds) Varieties of attention. Academic Press, New York, pp 29–61

Kahneman D, Treisman A, Gibbs BJ (1992) The reviewing of object files: object-specific integration of information. Cogn Psychol 24:174–219

Kenny DA, Kashy DA, Bolger N (1998) Data analysis in social psychology. In: Gilbert D, Fiske S, Lindzey G (eds) The handbook of social psychology, vol 1, 4th edn. McGraw-Hill, Boston, pp 233–265

Kilian A, Yaman S, von Fersen L, Güntürkün O (2003) A bottlenose dolphin discriminates visual stimuli differing in numerosity. Learn Behav 31(2):133–142

Kitchen DM, Cheney DL, Seyfarth RM (2004) Factors mediating inter-group encounters in savannah baboons (*Papio cynocephalus ursinus*). Behaviour 141:197–218

Libertus ME, Pruitt LB, Woldorff MG, Brannon EM (2009) Induced alpha-band oscillations reflect ratio-dependent number discrimination in the infant brain. J Cog Neurosci 21(12):2398–2406

MacKinnon DP, Warsi G, Dwyer JH (1995) A simulation study of mediated effect measures. Multivar Behav Res 30:41–62

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

Meck WH, Church RM (1983) A mode control model of counting and timing processes. J Exp Psychol Anim Behav Process 9:320–334

Mix K, Huttenlocher J, Levine S (2002) Multiple cues for quantification in infancy: is number one of them? Psychol Bull 128: 278–294

Pack AA, Herman LM, Roitblat HL (1991) Generalization of visual matching and delayed matching by a California sea lion (*Zalophus californianus*). Anim Learn Behav 19:37–48

Reichmuth Kastak C, Schusterman RJ (2002) Long term memory for concepts in a California sea lion. Anim Cogn 5:225–232

Schusterman RJ, Reichmuth Kastak C, Kastak D (2002) The cognitive sea lion: meaning and memory in the lab and in nature. In: Bekoff M, Allen C, Burghardt G (eds) The cognitive animal: empirical and theoretical perspectives on animal cognition. MIT Press, Cambridge, pp 217–228

Sepulveda M, Perez MJ, Sielfeld W, Oliva D, Duran LR, Rodriguez L, Araos V, Buscaglia M (2007) Operational interaction between South American sea lions (*Otaria flavescens*) and artisanal

(small-scale) fishing in Chile: results from interview surveys and on-board observations. Fish Res 83(2–3):332–340
Shettleworth S (2009) The evolution of comparative cognition: is the snark still a boojum? Behav Process 80:210–217
Shettleworth S (2010) Cognition, evolution, and behaviour. Oxford University Press, Oxford
Simon TJ, Hespos SJ, Rochat P (1995) Do infants understand arithmetic? A replication of Wynn (1992). Cogn Dev 10:253–269
Sobel ME (1982) Asymptotic confidence intervals for indirect effects in structural equation models. Sociol Methodol 13:290–312
Thomas RK (1992) Primates' conceptual use of number: ecological perspectives and psychological processes. In: Nishida T, McGrew WC, Marler P, Pickfbrd M, de Waal FBM (eds) Topics in primatology, vol. 1: human origins. Tokyo University Press, Tokyo, pp 305–314
Tomasello M, Call J (1997) Primate cognition. Oxford University Press, Oxford
Uller C (2008) Developmental and evolutionary considerations on numerical cognition: a review. J Evol Psychol 6(4):237–253
Uller C, Lewis J (2009) Horses (*Equus caballus*) select the greater of two quantities in small numerical contrasts. Anim Cogn 12(5):733–738
Uller C, Carey S, Huntley-Fenner G, Klatt L (1999) What representations might underlie infant numerical knowledge. Cogn Dev 14(1):1–36
Ward C, Smuts B (2007) Quantity-based judgments in the domestic dog (*Canis lupus familiaris*). Anim Cogn 10:71–80
Watts D, Mitani J (2002) Hunting behaviour of chimpanzees at Ngogo, Kibale National Park, Uganda. Int J Primatol 23:1–28
Wiese H (2003) Iconic and non-iconic stages in number development: the role of language. Trends Cogn Sci 7:385–390
Wilson M, Hauser M, Wrangham R (2007) Chimpanzees (*Pan troglodytes*) modify grouping and vocal behaviour in response to location-specific risk. Behaviour 144:1621–1653
Xu F, Carey S (1996) Infants' metaphysics: the case of numerical identity. Cogn Psychol 30:111–153