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

Daniel Hanus and Josep Call
Max Planck Institute for Evolutionary Anthropology

The authors examined quantity-based judgments for up to 10 items for simultaneous and sequential whole sets as well as for sequentially dropped items in chimpanzees (*Pan troglodytes*), gorillas (*Gorilla gorilla*), bonobos (*Pan paniscus*), and orangutans (*Pongo pygmaeus*). In Experiment 1, subjects had to choose the larger of 2 quantities presented in 2 separate dishes either simultaneously or 1 dish after the other. Representatives of all species were capable of selecting the larger of 2 quantities in both conditions, even when the quantities were large and the numerical distance between them was small. In Experiment 2, subjects had to select between the same food quantities sequentially dropped into 2 opaque cups so that none of the quantities were ever viewed as a whole. The authors found some evidence (albeit weaker) that subjects were able to select the larger quantity of items. Furthermore, the authors found no performance breakdown with the inclusion of certain quantities. Instead, the ratio between quantities was the best performance predictor. The authors conclude that quantity-based judgments rely on an analogical system, not a discrete object file model or perceptual estimation mechanism, such as subitizing.

Keywords: quantity estimation, relative numerousness, great apes, accumulator model, quantitative abilities

Many animals engage in quantitative judgments to deal with a variety of problems, including the maximization of the daily food intake or the quantitative estimation of opponent strength in aggressive encounters (e.g., McComb, Packer, & Pusey, 1994; Tomasello & Call, 1997). A key component for solving quantitative problems consists of estimating and comparing the quantities available in different sets. Many studies on quantity assessment have relied on the simultaneous or sequential presentation of discrete quantities, typically including less than seven items per pair (e.g., J. R. Anderson, Awazu, & Fujita, 2000; Boysen & Berntson, 1995; Call, 2000; Dooley & Gill, 1977; Silberberg & Fujita, 1996). These studies show that animals spontaneously select those sets with the larger quantities.

Much debate has revolved around the mechanisms that are responsible for this spontaneous performance. Initially, perceptual apprehension mechanisms, such as subitizing (Davis & Perusse, 1988) or prototype matching (Thomas, 1992), had been proposed to explain these findings. Subitizing can be described as a perceptual low-level process through which a certain number of items can just be "seen" without the need of counting or estimating it. Even though it is still under debate, the limit of that mechanism seems to lie between three and four items (Mandler & Shebo, 1982; Pylyshyn, 1989, 1994). However it needs to point out that subitiz-

ing in general is a rather descriptive than explanatory concept for quantitative perception (Terrell & Thomas, 1990). Prototype matching describes another process used for quantitative reasoning within small numerical ranges. The underlying idea is that perceived new stimuli are compared with abstract, idealized patterns that are stored in memory. Such patterns are called prototypes, and they are essential to establish numerousness concepts, for example, the concept of "twoness" (Thomas & Lorden, 1993). A stimulus array gets the label "two" only if the current image matches with that certain concept. This process appears to have an upper limit of seven or eight (Terrell & Thomas, 1990), perhaps more, although this has not been demonstrated.

However, these mechanisms cannot account for some of the data currently available because some subjects discriminated between quantities that fell outside the subitizing range (Dooley & Gill, 1977) and were presented under conditions of spatial segregation between sets (Rumbaugh, Savage-Rumbaugh, & Hegel, 1987). Moreover, apes can also select the larger of two quantities even if sets are presented sequentially so that the pairs are never simultaneously available (Beran, 2001; Call, 2000). This means that subjects have to encode the information about the quantity located in each dish and mentally compare them to select the larger amount, which pushes a purely perceptual mechanism to unbearable limits.

More important, some studies have also shown that subjects can cope with quantities that are presented not as a whole set, but item-by-item. Following Wynn's (1992) study with infants, Hauser, Carey, and Hauser (2000) tested rhesus macaques by showing them pieces of food being hidden behind a screen. Using the expectancy violation technique, they found that after the screen was lifted, the monkeys looked longer when the number of food

Daniel Hanus and Josep Call, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

Correspondence concerning this article should be addressed to Daniel Hanus, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103, Leipzig, Germany. E-mail: hanus@eva.mpg.de

that appeared did not match with the expected amount (e.g., two items went successively behind the screen, and one or three items appeared after removing the screen). Several authors have proposed the object file model introduced by Kahneman and Treisman (1984; see also Kahneman, Treisman, & Gibbs, 1992) and later refined by Uller and colleagues (Uller, Carey, & Hauser, 1997; Uller, Carey, Huntley-Fenner, & Klatt, 1999) and Simon, Hespos, and Rochat (1995) to account for these results. According to this model, young children and other animals construct a mental image of a given scene that provides quantitative information (Uller et al., 1997; Uller et al., 1999). In doing so they create one object file for each (countable) object and thus produce a one-to-one correspondence between "active" object files and the number of presented items (Hauser & Carey, 1998). Subjects are able to detect quantitative differences between two scenes or arrays by detecting the missing correspondence of the two representations. Because this system needs memory space for keeping the used objects files in an active state, the short-term storage load should be the limiting factor. Indeed, it has been argued that young infants and nonhuman animals have problems when dealing with quantities bigger than three or four for exactly that reason (Uller et al., 1999). In an experiment with semi-free-ranging rhesus monkeys, the subjects had to find the bigger of two amounts of apple slices that were placed in a container one at a time (Hauser et al., 2000). Hauser et al. found that the monkeys were successful for small quantity pairs (e.g., three vs. four) but failed as soon as they had to make comparisons based on more than three pieces per container (e.g., four vs. five). Results from a human infant study with a similar setting reached comparable conclusions (Feigenson, Carey, & Hauser, 2002). Twelve-month-old infants preferably visited locations where they saw the bigger quantity of crackers being dropped into two opaque containers. More important, they only were successful when each amount was smaller than four (e.g., succeeded on two vs. three but failed on two vs. four).

However, there are some studies that have reported tracking of more than four objects presented in a sequential fashion. Boysen and Berntson (1989) showed that a female chimpanzee was able to add the number of certain objects she had found at several locations in a room. She could also do the same with Arabic numerals. Recently, Pepperberg (2006) has also shown that the parrot Alex is capable of combining quantities that are presented discretely by means of symbolic labels. Clearly, these findings cannot be easily explained by invoking the object file model. However, it is unclear to what extent they are based on a process available to all members of a species and not just those that subjects had been explicitly trained to use such a symbolic system. Obviously, the use of such a symbolic system is important in itself, but it does not show whether animals can use other mechanisms to deal with larger quantities in the absence of training on the use of symbolic devices.

The accumulator model (Meck & Church, 1983) is a mechanism that has been proposed to deal with larger quantities in the absence of explicit training. According to this model, discrete and continuous quantities (e.g., event durations) are represented in principle in the same manner as mental magnitudes. These magnitudes are accumulated in a way that a certain magnitude represents a certain number, like a certain level of water in a measuring cup represents a certain volume. Following that logic, the amount of the magnitude in the accumulator is larger for greater numbers. Thus, the

represented magnitudes could be seen as an analog of the perceived discrete quantities (or event durations). It is important to note that this representation (magnitude) is rather noisy (Gallistel & Gelman, 2000) because the accuracy of discrimination decreases with increasing quantities (when the absolute distance is kept constant), but, importantly, it has no a priori limit.

This model could explain the ability of chimpanzees to discriminate between relatively large sets of items presented one-by-one in the absence of symbolic devices. Beran (2001) presented two chimpanzees with up to nine chocolate candies that were sequentially dropped into two opaque cups. Even if the total number for each cup was not dropped into the cup continuously, but instead during three intervals between which items were dropped into the other cup, the subjects selected significantly more often the side with the larger amount. In another study from the same laboratory, two language trained chimpanzees saw the sequential dropping of two different quantities into two opaque cups (up to 10 items per side) plus one open visible quantity (Beran, 2004). Both subjects were capable of detecting the largest of those three amounts by comparing the two nonvisible and the one visible set of food pieces. Beran and Beran (2004) also demonstrated chimpanzee capacity of extended memory in a quantity discrimination task. Here the four subjects were able to remember the location of the larger quantity even after a time delay of 20 min from initial presentation.

Although these results are important, the reduced sample size of these studies begs for replication with additional chimpanzees. Moreover, albeit the chimpanzees studied by Beran (2001) were not trained on this task, both had ample experience using symbolic devices and solving a variety of numerical problems. It is therefore unclear whether animals without such numerical training would also be able to solve this problem. The small number of subjects also means that there is little information about the parameters controlling the subjects' choices, which may affect the ability to identify those factors governing their choices. Distance and magnitude effects have been often reported in the literature (Brannon & Roitman, 2003; Dehaene, 1997), with ratio between quantities being one of the most commonly used variables (Beran, 2001; Boysen, Mukobi, & Berntson, 1999; Call, 2000; Huntley-Fenner & Cannon, 2000). However, other variables, such as the difference between quantities or the total quantity, may also explain some of the variance independently from ratio. Finally, it is unknown to what extent these findings may generalize to other great ape species, particularly to individuals without formal training in numerical tasks. There are no data on bonobo quantitative skills and comparatively very little on gorillas (U. S. Anderson et al., 2005) and orangutans (Call, 2000). Such comparisons are crucial to attempt to reconstruct the evolution of cognitive skills.

In the current study, we investigated the ability of all great ape species to estimate, compare, and select the larger of two sets of quantities presented in three different ways: whole sets presented simultaneously or sequentially (Experiment 1), and sets presented item-by-item (Experiment 2). Experiment 1 borrowed the method used by Call (2000), which consisted of presenting pairs of whole quantity sets either simultaneously so that the quantities could be directly compared or sequentially so that no such direct comparison was possible. Experiment 2 borrowed the method used by Beran (2001), which consisted of dropping the items one-by-one into two opaque cups, so that none of the quantities were ever

viewed in their totality and could not be directly compared. Including all three types of presentation allowed us to compare them directly. Moreover, we tested all great ape species with the exact same procedure. All apes that participated in this study had never before participated in studies assessing quantitative abilities. To explore the limits in each of these two experiments, we varied the quantity sets in both experiments between 0 and 10 items, which represent the upper limits used in previous ape studies (Beran, 2001; Dooley & Gill, 1977). We investigated the effect of the ratio between quantities, the difference between quantities, and the total quantities used in the subjects' choices for each type of presentation. A predominant influence of the factor ratio on apes' performance would support an analog representational mechanism as suggested by the accumulator model.

Experiment 1: Continuous Presentation of Rewards (Dishes)

Method

Subjects

We tested 4 bonobos (*Pan paniscus*), 12 chimpanzees (*Pan troglodytes*), 7 gorillas (*Gorilla gorilla*), and 7 orangutans (*Pongo pygmaeus*) housed at the Wolfgang Köhler Primate Research Center in the Leipzig-Zoo, Germany. There were 20 females and 10 males ranging from 5 to 31 years of age. All male bonobos and all the adult chimpanzees were nursery reared, whereas all other subjects were mother reared. All subjects lived in social groups of various sizes, with access to indoor and outdoor areas. Prior to this experiment, subjects had participated in a number of cognitive studies that involved selecting one of several containers to get the reward. One of those studies investigated Piagetian conservation, and subjects had to select between two different liquid quantities presented in containers that differed in shape and amount (Suda & Call, 2004, 2006). Subjects were individually tested in their indoor cages and were not food or water deprived. Table 1 provides detailed information about each subject, including its participation in the various tests.

Materials

Two identical white dishes (7 cm in diameter and 1 cm deep) with their corresponding lids were used to present sets of dark brown food pellets of approximately 1 cm in length. The number of pellets inside each dish varied between 0 and 10. Dishes were presented next to each other on a wooden platform (80 × 37 cm) with a distance of 20 cm from the center of the platform to the center of each dish.

Procedure

Subjects sat across from the experimenter separated by a Plexiglas window that had three small holes on its bottom through which subjects could indicate their choices by touching the dish that contained the contents they wanted to receive. The experimenter filled each of the dishes with a certain amount of pellets out of the subject's view, covered each dish with the lid, placed the dishes on the platform, and administered one of the two following presentation conditions:

Table 1

Apes That Participated in Experiment 1 (Test 1: Small Quantities; Test 2: Large Quantities) and Experiment 2 (Test 3: Small Quantities; Test 4: Large Quantities)

Subject	Species	Sex	Age (years)	Test participation
Joey	*Pan paniscus*	M	21	1, 2
Kuno	*Pan paniscus*	M	7	1, 2, 3, 4
Limbuko	*Pan paniscus*	M	8	1, 2, 3, 4
Ulindi	*Pan paniscus*	F	10	1, 2
Alexandra	*Pan troglodytes*	F	5	2
Annett	*Pan troglodytes*	F	5	1, 2
Brent	*Pan troglodytes*	M	5	1, 2
Dorien	*Pan troglodytes*	F	24	1, 2, 3, 4
Fifi	*Pan troglodytes*	F	11	1, 2, 3
Fraukje	*Pan troglodytes*	F	28	2, 3, 4
Frodo	*Pan troglodytes*	M	10	1, 2, 3, 4
Gertruida	*Pan troglodytes*	F	11	1, 2, 3, 4
Jahaga	*Pan troglodytes*	F	11	1
Natascha	*Pan troglodytes*	F	24	3
Patrick	*Pan troglodytes*	M	7	1, 2, 3
Pia	*Pan troglodytes*	F	5	1, 2, 3
Sandra	*Pan troglodytes*	F	11	1, 2, 3, 4
Bebe	*Gorilla gorilla*	F	25	1, 2, 3, 4
Gorgo	*Gorilla gorilla*	M	23	2, 3
N'Diki	*Gorilla gorilla*	F	27	1, 2, 3, 4
N'Kwango	*Gorilla gorilla*	M	6	1, 3, 4
Ruby	*Gorilla gorilla*	F	6	1, 2, 3, 4
Viringika	*Gorilla gorilla*	F	9	2, 3, 4
Vizuri	*Gorilla gorilla*	F	9	1
Bimbo	*Pongo pygmaeus*	M	23	1, 2, 3
Dokana	*Pongo pygmaeus*	F	15	2, 3, 4
Dunja	*Pongo pygmaeus*	F	31	1, 2, 3, 4
Padana	*Pongo pygmaeus*	F	7	1, 2, 3, 4
Pini	*Pongo pygmaeus*	F	17	1, 2, 3, 4
Toba	*Pongo pygmaeus*	F	10	1, 2, 3, 4
Walter	*Pongo pygmaeus*	M	15	1, 2

Simultaneous. The experimenter removed both lids simultaneously so that the subject could see the contents of each dish. As soon as the subject had watched the contents for 3–5 s, both dishes were moved within the reach of the subject so that she could make her choice.

Successive. The experimenter removed the lid from the left dish while the right dish remained covered. When the subject had seen the contents of the open dish for 3–5 s, that dish was covered again with its lid. The same operation was repeated with the dish on the right side before both (covered) dishes were moved within the subject's reach to allow her to select one of them.

All subjects experienced both presentation conditions in each of two tests that differed only in the number of pellets located inside each dish (small vs. large sets). Subjects received the small quantity test first and the large quantity test second. If subjects lost interest during presentation and left the testing station, the trial was canceled and restarted.

In the small quantities test, we presented all possible combinations from zero to six pellets, except that the "quantity of zero" was only paired with the quantity of one (i.e., 0:1, 1:2, 1:3, 1:4, 1:5, 1:6, 2:3, 2:4, 2:5, 2:6, 3:4, 3:5, 3:6, 4:5, 4:6, 5:6). Each combination was presented four times for each of the two presentation conditions resulting in 128 trials in total (16 combinations × 4 trials × 2 presentation conditions). Both conditions were presented in a

mixed fashion with each condition appearing at most two times in a row and were counterbalanced with the larger amount never appearing more than two times in a row on the same side. Depending on individual attention span, each subject received between 6 and 12 trials per session, culminating in 12–21 testing days per subject in total.

In the larger quantities test, we increased the number of possible items in each dish to a maximum of 10 and presented 18 selected pairs. Five pairs consisted of combinations that subjects had already experienced in the smaller quantities test (i.e., 1:2, 2:3, 3:6, 4:6, 5:6), whereas the remaining 13 pairs were novel (i.e., 3:9, 4:8, 4:10, 5:8, 5:9, 6:8, 6:9, 6:10, 7:8, 7:9, 7:10, 8:10, 9:10). We selected those pairs that allowed us to investigate the effect of the total amount of pellets available, as well as the difference and the ratio between quantities. Because some of these variables are highly correlated, we selected certain pairs that coincided in some values of these variables but differed in others. For instance, the pairs 1:2, 2:4, and 4:8 shared the same ratio (0.5) but differed in the difference between the pair quantities (1 vs. 2 vs. 4). Each of the 18 pairs was presented twice for each of the two presentation conditions resulting in 72 trials in total. Besides the changes in the number of items and the number of trials administered, the procedure was exactly the same as in the smaller quantities test.

Data Scoring and Analysis

All trials were videotaped. Our dependent variable was the percentage of trials in which subjects selected the larger quantity of the pair. Subjects indicated their choices by touching one of the dishes with their fingers (or very rarely with their tongue). They received the contents of the dish that they touched first. Trials were repeated when the subject touched both dishes simultaneously. A second observer (Josep Call) scored 20% of the sessions to assess interobserver reliability. Interobserver reliability was excellent (Cohen's κ = .99).

We investigated the effect of presentation condition and species on our dependent variable using nonparametric statistics (Wilcoxon, Kruskall–Wallis, and Mann–Whitney tests) because our data did not meet the supposition of homogeneity of variance. We used stepwise multiple regression to investigate how three variables (total quantity: smaller quantity + larger quantity; ratio: smaller quantity/larger quantity; difference: larger quantity − smaller quantity) affected our dependent variable. We also assessed whether subjects performed above chance levels both at the group (one-sample t test) and individual levels (Binomial test).

Results

Figure 1a presents the mean percentage of correct trials in the low-quantity pairs in each condition for each species. At the group level, all species performed above chance in both conditions (one-sample t test: p < .05 in all cases). There were no species differences in the simultaneous (Kruskall–Wallis test = 7.07, df = 3, p = .07, n = 25) or sequential conditions (Kruskall–Wallis test = 4.44, df = 3, p = .22, n = 25). Additionally, there were no significant differences between the simultaneous and the sequential condition (Wilcoxon test: z = 0.31, p = .75). Individual analyses also indicated that all subjects were above chance (Binomial test: p < .001).

Figure 1b presents the mean percentage of correct trials in the high-quantity pairs in each condition for each species. At the group level, all species performed above chance in both conditions (one-sample t test: p < .05 in all cases). There were no species differences in the simultaneous condition (Kruskall–Wallis test = 3.41, df = 3, p = .33, n = 27). In contrast, species differed significantly in the sequential condition (Kruskall–Wallis test = 7.80, df = 3, p = .05, n = 27). However, Bonferroni–Holm post hoc pairwise comparisons failed to confirm such differences (p > .05 in all cases). Additionally, there were no significant differences between the simultaneous and the sequential condition (Wilcoxon test: z = 0.07, p = .95). Individual analyses also indicated that all subjects except 2 gorillas (Bebe and Viringika) and 1 orangutan (Dokana) were above chance (Binomial test: p < .05).

We also analyzed whether subjects' performance improved with further testing by comparing the overall score for the pairs that appeared in both the low- and high-quantity tests. We found no evidence of improvement across tests (Wilcoxon test: z = 0.47, p = .64).

In our next analysis, we combined the data from the low- and the high-quantity tests to investigate the variables that affected the subjects' performance. A multiple regression with ratio, difference, and total quantities as independent variables and the percentage of correct trials as the dependent variable revealed that a model composed by the ratio and difference produced the best fit, R = 0.943, $F(2, 26)$ = 103.48, p < .001. Ratio alone explained 81% of the variance in performance. Subjects' performance increased as a function of decreasing ratios and increasing differences between quantities ($\hat{y}$ = −0.301 × Ratio + 0.02 × Difference + 0.919) (see Figure 2). Analyzing all four species separately also revealed that ratio was the factor that accounted for most of the variance in each species.

Overall, subjects performed above chance in all pairs (one-sample t test): $t(26)$ = 3.30, p < .005 in all cases, including those with the largest ratios and smallest difference between quantities (one-sample t test): Pair 8/10: $t(26)$ = 3.24, p = .003; Pair 5/6: $t(26)$ = 3.86, p = .001; Pair 7/8: $t(26)$ = 3.32, p = .003; Pair 9/10: $t(26)$ = 3.68, p = .001.

Discussion

Overall, all species selected the larger quantity in all possible pairs regardless of the type of presentation (simultaneous or sequential). There were no significant differences between species, and all subjects except 2 gorillas and 1 orangutan performed above chance in both the low-quantity and high-quantity tests. The ratio of the pairs defined as the smaller quantity divided by the larger quantity combined with the difference between quantities explained most of the variability in our data. In particular, high ratios combined with small differences between quantities constituted the most challenging pairs.

These results confirm previous data showing that apes can discriminate between quantities even when they were prevented from seeing both quantities simultaneously. Moreover, they permit us to extend previous results in several directions. First, bonobos and gorillas performed at comparable levels to chimpanzees and orangutans. Second, apes can make these discriminations even with very high ratios (i.e., >0.80) and relatively large quantities (i.e., 9 vs. 10)—something that had been only previously shown in
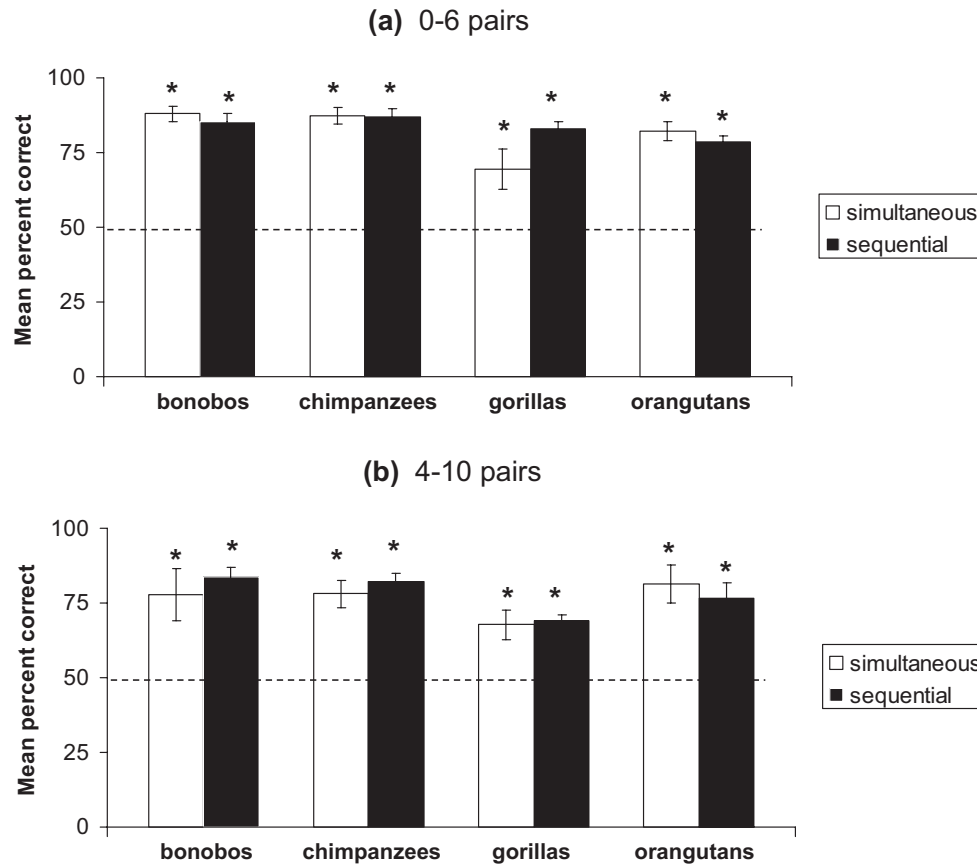
(a) 0-6 pairs

(b) 4-10 pairs

Figure 1. Percentage of trials in which subjects chose the larger quantity for each species as a function of the experimental condition (simultaneous and sequential type of presentation). (a) Small quantity pairs: 0–6 pellets; (b) large quantity pairs: 4–10 pellets. $^*p < .05$. Error bars depict the standard errors of the means.
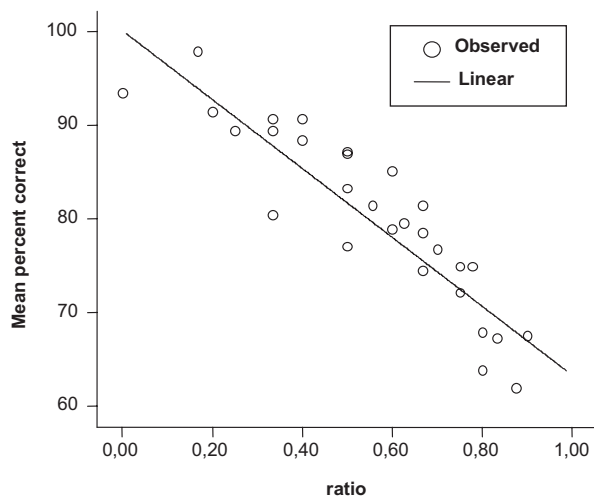

Figure 2. Percentage of trials in which subjects chose the larger quantity as a function of the ratio (small/large) of the two quantities.

a chimpanzee (but see Brannon & Terrace, 2000, for comparable performance of rhesus macaques in an ordinal judgment task). Third, the combined effect of ratio and difference between pairs helped to explain the variability to a very high degree. In the next experiment, we investigated whether subjects would be able to maintain their good performance when arrays were introduced item-by-item as opposed to all at once.

Experiment 2: Discontinuous Reward Presentation (Cups)

Method

Subjects

We tested all apes that had participated in Experiment 1 except 2 bonobos, 4 chimpanzees, 1 gorilla, and 1 orangutan that were either not available at the time of testing (transfer to another zoo) or did not show enough attention to the new and longer lasting procedure. We also tested a chimpanzee that had not participated in the previous experiment but that belonged to the same group of chimpanzees (see Table 1 for details). Thus, our sample included 2 bonobos, 9 chimpanzees, 6 gorillas, and 6 orangutans. There were 17 females and 6 males ranging from 5 to 32 years of age. Subjects were individually tested in their indoor cages and were not food or water deprived.

Materials

We used the same food quantities and presentation format as in Experiment 1 except that we replaced the pair of dishes for a pair of identical white cups (8 cm in diameter and 17 cm in height) with their corresponding lids.

Procedure

We used the same basic procedure as in Experiment 1 except for the presentation format of the pellets and the number of trials for each pair. Instead of showing the subject the pellets resting on the bottom of a dish, we dropped them one-by-one into the cups. Prior to the dropping of the pellets, the experimenter gave the subject the opportunity to look into the cups to demonstrate that they were empty. After that, both cups were placed on the platform in an upright position. The food pellets were then dropped into the cups successively, starting with the left cup and then the right cup. In doing so, subjects could only see one item at a time appearing between the thumb and index finger of the experimenter. Once the last pellet was dropped into the right cup, both cups were moved within reach of the subject to allow the subject to make her choice. For some subjects that tried to look inside the cup (even though the angle was very awkward), we also placed a lid on top of each cup at the end of the procedure to block their view of the inside of the cups. If subjects lost interest during presentation and left the testing station, the trial was canceled and restarted.

As in Experiment 1, all subjects received two consecutive tests that differed only in the number of pellets located inside each cup. Subjects received the small quantity test first and the large quantity test second. In the small quantities test, we presented all possible combinations from zero to six pellets, except that the "quantity of zero" was only paired with the quantity of one, thus duplicating the quantities used in Experiment 1. Each of the 16 combinations was presented four times resulting in 64 trials in total. In the larger quantities test, we followed the same presentation procedure as above but increased the number of possible items in each dish to a maximum of 10 pellets and presented the same selected pairs that we had used in Experiment 1. Each of the 18 pairs was presented twice for each of the two presentation conditions resulting in 72 trials in total. Trials were counterbalanced with the larger amount never appearing more than two times in a row on the same side. Depending on the individual's attention span, each subject received between 6 and 12 trials per session, culminating in 7–11 testing days per subject in total.

Data Scoring and Analysis

We used the same scoring method and analyses as in Experiment 1. Interobserver reliability was excellent (Cohen's $\kappa = 1.0$).

Results

Figure 3a presents the mean percentage of correct trials in the low-quantity pairs for each species. At the group level, all species selected the larger quantity above chance (one-sample t test: $p <$ .05 in all cases) except bonobos (one-sample t test: $p = .058$). However, there were no species differences in performance (Kruskall–Wallis test $= 0.71$, $df = 3$, $p = .87$, $n = 23$). Individual

analyses indicated that 4 chimpanzees and 2 orangutans were above chance (Binomial test: $p < .05$ in all cases).

Figure 3b presents the mean percentage of correct trials in the high-quantity pairs for each species. None of the species performed above chance (one-sample t test: $p > .05$ in all cases). As in the pervious experiments, there were no species differences (Kruskall–Wallis test $= 1.27$, $df = 3$, $p = .74$, $n = 17$). Individual analyses indicated that only 1 chimpanzee and 1 orangutan were above chance (Binomial test: $p < .05$ in both cases).

We also analyzed whether subjects' performance improved with further testing by comparing the overall score for the pairs that appeared in both the low- and the high-quantity tests. We found no evidence of improvement across tests (Wilcoxon test: $z = 0.45$, $p = .96$).

In our next analysis, we combined the data from the low- and the high-quantity tests to investigate the variables that affected the subjects' performance. A multiple regression with ratio, difference, and total quantities as independent variables and the percentage of correct trials as the dependent variable revealed that a model composed by the ratio and the square of the ratio produced the best fit, $R = 0.726$, $F(1, 27) = 14.46$, $p < .001$. Ratio alone explained 38% of the variance in performance. Subjects' performance increased as a function of decreasing ratios between quantities ($\hat{y} = 0.494 \times$ Ratio $2 - 0.687 \times$ Ratio $+ 0.807$) (see Figure 4).

Overall, subjects performed above chance (one-sample t test) in the following 11 pairs: 0/1: $p < .001$; 1/2: $p = .029$; 1/3: $p = .006$; 1/4: $p = .004$; 1/5: $p < .001$; 2/5: $p = .047$; 4/5: $p = .015$; 1/6: $p < .001$; 2/6: $p < .001$; 5/6: $p = .045$; 5/8: $p = .004$.

Finally, we found no correlation between the performance in Experiments 1 and 2 ($r = .343$, $p = .252$, $n = 13$).

Discussion

Overall, all species selected the larger quantity in the low-quantity test but not in the high-quantity test. There were no significant differences between species, although bonobos failed to perform above chance. However, only 4 chimpanzees and 2 orangutans were above chance in the low-quantity test, whereas this number dropped to 1 chimpanzee and 1 orangutan in the high-quantity test. The ratio between quantities was the variable that explained best the data variability, but it did so much less than in Experiment 1.

These results show that some chimpanzees and orangutans can discriminate between pairs of quantities even when the items in each pair are presented one-by-one, thus replicating Beran's (2001; Beran & Beran, 2004) results on chimpanzees and extending them to orangutans. However, the chimpanzees in Beran and Beran's (2004) study outperformed even our best subjects. This cannot be explained as a result of a lack of understanding regarding the procedure because they succeeded in some pairs, particularly those with small ratios between quantities. It is very likely that the high level of experimental experience of Beran's two individuals (Lana and Sherman) in language and quantity-related tasks may have substantially contributed to their high-level performance.

Finally, our study also demonstrated that not only chimpanzees, but also orangutans, showed some ability to solve this problem. We find it interesting that these two species are the most frugivorous apes. One may speculate that these species experienced higher selective pressure compared with the other apes, but it must
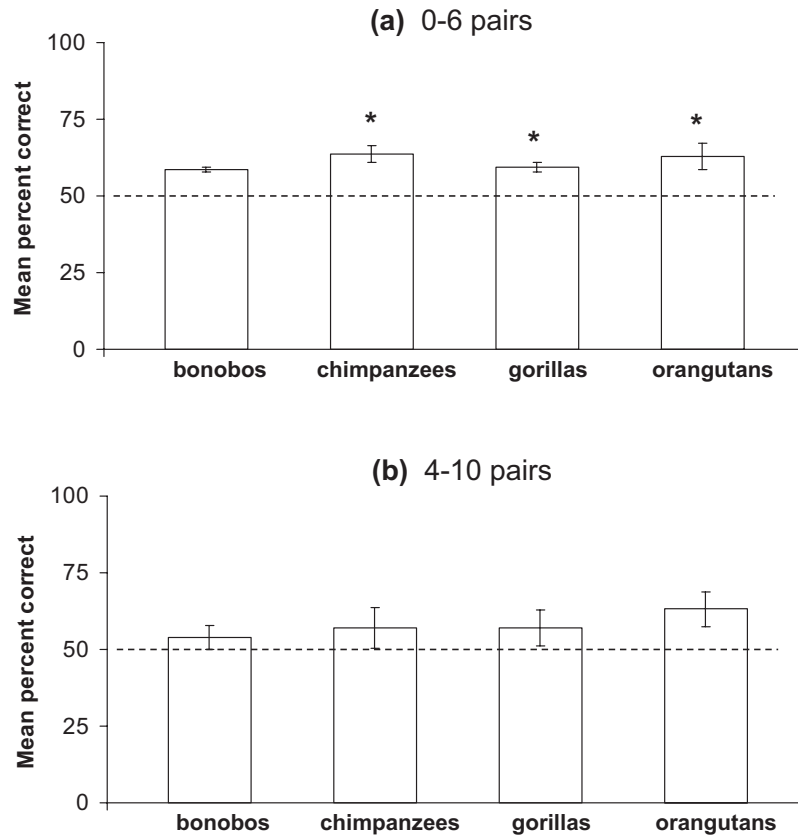
(a) 0–6 pairs

(b) 4–10 pairs

Figure 3. Percentage of trials in which subjects chose the larger quantity for each species. (a) Small quantity pairs: 0–6 pellets; (b) large quantity pairs: 4–10 pellets. $^*p < .05$. Error bars depict the standard errors of the means.

Figure 4. Percentage of trials in which subjects chose the larger quantity as a function of the ratio (small/large) of the two quantities.

be emphasized that our data did not show clear-cut species differences.

General Discussion

Apes selected the larger of two quantities presented as pairs of whole sets either simultaneously or sequentially. Some apes also solved the problem when the quantities were presented item-by-item but mostly with small sets. We found no significant species differences although there were important interindividual differences, especially in the item-by-item presentation. Ratio between quantities was the best predictor of performance both when items were presented as whole sets or one-by-one. Performance dropped as a function of increasing ratio between quantities, although its exact function and its predictive power depended on the type of presentation.

Thus, our results confirm previous studies on relative quantity discrimination of whole sets in chimpanzees and orangutans (Beran, 2001; Call, 2000) and extend them to larger quantities that had not been tested before, particularly in the case of orangutans. Moreover, our results also show that both gorillas and bonobos are capable of solving this problem. Our data on the item-by-item presentation also replicate the findings with chimpanzees (Beran, 2001), albeit with less robust results. Bonobos, however, failed to perform significantly above chance, although this may have been
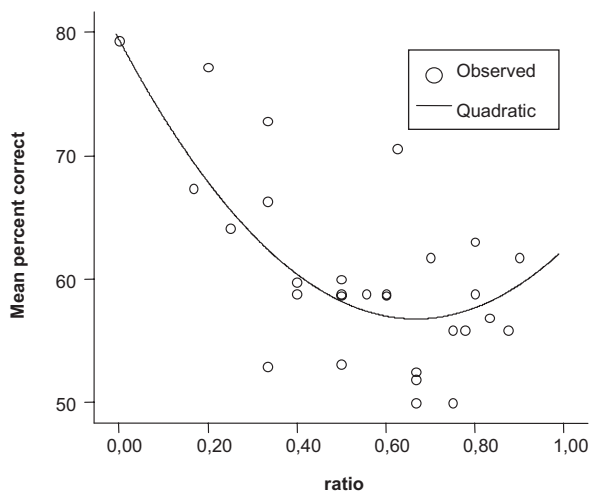
partly a result of our small sample size. Beran (2001) reported an overall accuracy of 78% for 2 chimpanzees, whereas the apes in the current study were correct on average in only 62% of the trials in the low-quantity test and 59% in the high-quantity test. A comparison with the data reported by Beran (2004), which used a very similar method to the current study, is also informative. Chimpanzees in the current study chose the larger quantity in 62% of the cases, whereas Lana's and Sherman's average success rate was 82%. However, the apes in the current study performed slightly better than rhesus macaques (58% vs. 50%) tested by Hauser et al. (2000) for those nine pairs of quantities that were common in both studies.

Our scores in the item-by-item presentation also contrast markedly with the performance in the whole set presentation (Experiment 1) in which subjects were correct on average in 83% of the trials in the low-quantity test and 77% in the high-quantity test. A comparison based on the performance of individuals highlights the differences between the two types of presentation even further. All subjects were above chance in the whole set presentation for the small-quantity test compared with 26% of the subjects in the item-by-item presentation. Similarly, 90% of the subjects were above chance in the whole set presentation for the high-quantity test, whereas only 9% were above chance in the item-by-item presentation. One major contributing factor to these differences were the higher attentional demands required by the item-by-item presentation, especially for the larger quantities. Here the longer lasting procedure most likely created a motivational decline and more highly desired food might have attenuated this effect.

Another main issue of the current study was the question of species differences. None of the tested four great ape species performed as a group significantly better than any other, and all became more error-prone with increasing ratios between quantities in both experiments, not the total number of items involved. Thus, in contrast to the rhesus macaques in Hauser et al.'s (2000) study, here we could not appreciate any clear performance breakdown for quantity discriminations that went beyond three or four items. Even in the item-by-item presentation, our subjects were correct on pairs located beyond the range predicted by the object file model (e.g., 5 vs. 6 and 5 vs. 8). Instead, here performance decreased continuously, with ratio being the best predictor of performance in our data. Its inclusion in a model accounted for 81% of the variability in the whole set presentation and 38% for the item-by-item presentation. Therefore, the current data can be explained without hypothesizing different mechanisms for processing small and large quantities as claimed by the subitizing or the object file model.

Unlike previous studies, we found that the difference between quantities also explained some of the variability independently from the ratio between quantities. Although pairs of quantities with similar ratios produced similar performances, which suggests that the subjects' choices fit Weber's law, a greater difference between the quantities in each pair also resulted in a slight improvement in performance (about 2% for each additional difference point between quantities in the pairs) in the whole presentation setup. Taken together, these results mean that not only there was no limit in the quantities that apes could process but that their responses followed an analogical magnitude system. This strongly suggests that the accumulator model is a good candidate to explain the quantitative judgments in the current study.

One final word of caution is necessary. This study was not aimed at assessing numerical competence but the ability to assess and compare quantities under different presentation regimes. Obviously, we cannot extract any conclusions regarding number appreciation because factors, such as surface area (Experiment 1) or dropping duration (Experiment 2), are confounded with number. We chose not to control for such nonnumerical cues because our aim was to simulate problems that subjects may encounter in nature, such as fruits on a tree or conspecifics walking out of sight one-by-one. Even if subjects were using the timing of presentation or the area occupied by the food as the main cues, it still meant that apes had to encode and mentally compare sets of quantities (of mass or time).

In conclusion, we found that naïve subjects were capable of solving the problem of spatial and temporal discontinuity between food items to net the larger of two quantities, although whole set presentations were much easier than item-by-item presentations. Ratio between quantities appeared as the best predictor of performance regardless of the type of presentation.

References

Anderson, J. R., Awazu, S., & Fujita, K. (2000). Can squirrel monkeys (*Saimiri sciureus*) learn self-control? A study using food array selection tests and reverse-reward contingency. *Journal of Experimental Psychology: Animal Behavior Processes, 26,* 87–97.

Anderson, U. S., Stoinski, T. S., Bloomsmith, M. A., Marr, M. J., Smith, A. D., & Maple, T. L. (2005). Relative numerousness judgment and summation in young and old Western lowland gorillas. *Journal of Comparative Psychology, 119,* 285–295.

Beran, M. J. (2001). Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115,* 181–191.

Beran, M. J. (2004). Chimpanzees (*Pan troglodytes*) respond to nonvisible sets after one-by-one addition and removal of items. *Journal of Comparative Psychology, 118,* 25–36.

Beran, M. J., & Beran, M. M. (2004). Chimpanzees remember the results of one-by-one addition of food items to sets over extended time periods. *Psychological Science, 15,* 94–99.

Boysen, S. T., & Berntson, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology, 103,* 23–31.

Boysen, S. T., & Berntson, G. G. (1995). Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 82–86.

Boysen, S. T., Mukobi, K. L., & Berntson, G. G. (1999). Overcoming response bias using symbolic representations of number by chimpanzees (*Pan troglodytes*). *Animal Learning & Behavior, 27,* 229–235.

Brannon, E. M., & Roitman, J. D. (2003). Nonverbal representations of time and number in animals and human infants. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 143–182). Boca Raton, FL: CRC Press.

Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1–9 by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes, 26,* 31–49.

Call, J. (2000). Estimating and operating on discrete quantities in orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 114,* 136–147.

Davis, H., & Perusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences, 11,* 561–615.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* New York: Oxford University Press.

Dooley, G. B., & Gill, T. V. (1977). Acquisition and use of mathematical skills by a linguistic chimpanzee. In D. M. Rumbaugh (Ed.), *Language learning by a chimpanzee: The Lana project* (pp. 247–260). New York: Academic Press.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13,* 150–156.

Gallistel, C., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4,* 59–65.

Hauser, M. D., & Carey, S. (1998). Building a cognitive creature from a set of primitives: Evolutionary and developmental insights. In D. Cummins & C. Allen (Eds.), *The evolution of mind* (pp. 51–106). Oxford, England: Oxford University Press.

Hauser, M. D., Carey, S., & Hauser, L. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London Series B: Biological Sciences, 267,* 829–833.

Huntley-Fenner, G., & Cannon, E. (2000). Preschoolers' magnitude comparisons are mediated by a preverbal analog mechanism. *Psychological Science, 11,* 147–152.

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

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111,* 1–22.

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

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 320–334.

Pepperberg, I. M. (2006). Grey parrot (*Psittacus erithacus*) numerical abilities: Addition and further experiments on a zero-like concept. *Journal of Comparative Psychology, 120,* 1–11.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32,* 65–97.

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50,* 363–384.

Rumbaugh, D. M., Savage-Rumbaugh, S., & Hegel, M. T. (1987). Summation in the chimpanzee (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 13,* 107–115.

Silberberg, A., & Fujita, K. (1996). Pointing at smaller food amounts in an analogue of Boysen and Berntson's (1995) procedure. *Journal of the Experimental Analysis of Behavior, 66,* 143–147.

Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development, 10,* 253–269.

Suda, C., & Call, J. (2004). Piagetian liquid conservation in the great apes (*Pan paniscus, Pan troglodytes,* and *Pongo pygmaeus*). *Journal of Comparative Psychology, 118,* 265–279.

Suda, C., & Call, J. (2006). What does an intermediate success rate mean? An analysis of a Piagetian liquid conservation task in the great apes. *Cognition, 99,* 53–71.

Terrell, D. F., & Thomas, R. K. (1990). Number-related discrimination and summation by squirrel monkeys (*Saimiri sciureus sciureus* and *S. boliviensus boliviensus*) on the basis of the number of sides of polygons. *Journal of Comparative Psychology, 104,* 238–247.

Thomas, R. K. (1992, July). *Primates' conceptual use of number: Ecological perspectives and psychological processes.* Paper presented at the 13th Congress of the International Primatological Society, Tokyo, Japan.

Thomas, R. K., & Lorden, R. B. (1993). Numerical competence in animals: A conservative view. In S. T. Boysen & E. Capaldi (Eds.), *The development of numerical competence: Animal and human models comparative cognition and neuroscience* (pp. 127–147). Hillsdale, NJ: Erlbaum.

Tomasello, M., & Call, J. (1997). *Primate cognition.* London: Oxford University Press.

Uller, C., Carey, S., & Hauser, M. D. (1997, November). *Is language needed for constructing sortal concepts? A study with nonhuman primates.* Paper presented at the 21st annual Boston University conference on language development, Somerville, MA.

Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development, 14,* 1–36.

Wynn, K. (1992, August 27). Addition and subtraction by human infants. *Nature, 358,* 749–750.