Comparative Studies of Numerical Cognition in Nonhuman Primates:
From Numerical Comparison to Arithmetic

by

Sarah Mychal Jones

Department of Psychology and Neuroscience
Duke University

Date:__________________________

Approved:

___________________________
Elizabeth M. Brannon, Supervisor

___________________________
Amy S. Joh

___________________________
Stephen R. Mitroff

___________________________
Michael L. Platt

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School
of Duke University

2012
ABSTRACT

Comparative Studies of Numerical Cognition in Nonhuman Primates: From Numerical Comparison to Arithmetic

by

Sarah Mychal Jones

Department of Psychology and Neuroscience
Duke University

Date:_______________________

Approved:

___________________________
Elizabeth M. Brannon, Supervisor

___________________________
Amy S. Joh

___________________________
Stephen R. Mitroff

___________________________
Michael L. Platt

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology and Neuroscience in the Graduate School of Duke University

2012
Abstract

There is a long-standing claim that humans and nonhuman primates share an evolutionarily ancient system of nonverbal number representation. By and large, the focus in the field has been on providing existence proofs of numerical competence in wide-ranging taxa or using individual species as models for comparisons with humans. Recent findings in numerical cognition have suggested that evidence for approximate numerical abilities in nonhuman species may indicate that humans and animals share a cognitive system for representing numerosities nonverbally. To date, little is known about the contextual and quantitative limits of that system, or how those limits differ between species. The studies presented here take a comparative, behavioral approach to characterizing species differences and similarities in the approximate number system, and the contexts that affect that system. Collectively, this set of studies provides evidence that the approximate number system evolved in primates as a malleable system in which numerical representations are accessed spontaneously and improved through training. Despite the sensitivity of the system to experience and context individual differences in sensitivity are greater than species differences suggesting that the selective pressures that constrained its evolution were early and general and that species variation in social group size and diet have less influence on the ANS. Finally my studies indicate that the ANS supports approximate arithmetic and is consistent
with the idea that ANS representations evolved to allow animals to calculate the world around them.
Contents

Abstract........................................................................................................................................iv

List of Tables..................................................................................................................................xii

List of Figures..................................................................................................................................xiii

1. Introduction........................................................................................................................................1

1.1 Number: A case study for examining abstract thought without language ..................1

1.2 A shared system of approximate numerical representation........................................3

1.2.1 Ratio Dependence .................................................................................................................4

1.2.2 The Semantic Congruity Effect – another hallmark of the approximate number system .................................................................10

1.3 Approximate number system and food choice tasks .......................................................11

1.4 Do animals use number in the wild? ...............................................................................13

1.5 Comparative Cognition and the case for studying prosimians ..................................17

1.6 Two systems view ..................................................................................................................19

1.6.1 Object File System ..............................................................................................................19

1.6.2 Developmental literature ..................................................................................................21

1.6.3 Evidence from animal literature .........................................................................................22

1.6.4 Factors that may influence the recruitment of the ANS vs. object files ..................24

1.7 Last resort hypothesis ..............................................................................................................26

1.8 Mental arithmetic in nonhumans.........................................................................................28

1.8.1 Ordering ...............................................................................................................................28

1.8.2 Addition ...............................................................................................................................30
1.8.3 Subtraction ........................................................................................................... 32
1.8.4 Multiplication ....................................................................................................... 34
1.9 Research Approach ............................................................................................... 36


2.1 Experiment 1A ......................................................................................................... 44
2.1.1 Methods ............................................................................................................... 44
  2.1.1.1 Subjects ........................................................................................................... 44
  2.1.1.2 Apparatus ....................................................................................................... 44
  2.1.1.3 Task ................................................................................................................ 45
  2.1.1.4 Stimuli ............................................................................................................ 46
  2.1.1.5 Procedure ...................................................................................................... 47
  2.1.1.6 Analysis ......................................................................................................... 47
2.1.2 Results .................................................................................................................. 49
  2.1.2.1 Training ......................................................................................................... 49
  2.1.2.2 Testing .......................................................................................................... 50
  2.1.2.3 Ratio-dependent performance ...................................................................... 51
  2.1.2.4 Quantifying numerical sensitivity ................................................................. 52
  2.1.2.5 Surface area .................................................................................................. 56
  2.1.2.6 The effect of training on numerical acuity ..................................................... 58
2.2 Experiment 1B ......................................................................................................... 58
  2.2.1 Methods ............................................................................................................ 58
3.1.1.1 Subjects and apparatus ................................................................. 80
3.1.1.2 Task and Procedure ................................................................. 81
3.1.1.3 Prior Training............................................................................... 83
3.1.1.4 Testing......................................................................................... 83
3.1.1.5 Stimuli......................................................................................... 83
3.1.2 Results............................................................................................. 84
3.1.2.1 Accuracy .................................................................................... 84
3.1.2.2 Semantic Congruity Effect............................................................ 85
3.1.2.3 Context Effect............................................................................ 87
3.1.2.4 Timing of contextual shift ............................................................ 89
3.1.3 Discussion....................................................................................... 91
  4.1 Experiment 1 .................................................................................... 105
  4.1.1 Methods ..................................................................................... 105
    4.1.1.1 Subjects ................................................................................ 105
    4.1.1.2 Procedure.............................................................................. 106
  4.1.2 Results.......................................................................................... 108
  4.2 Experiment 2 .................................................................................. 110
  4.2.1 Methods ..................................................................................... 111
    4.2.1.1 Subjects ................................................................................ 111
    4.2.1.2 Procedure.............................................................................. 111
4.2.2 Results ......................................................................................................................... 111
4.3 Experiment 3 .................................................................................................................... 112
4.3.1 Methods ....................................................................................................................... 113
  4.3.1.1 Subjects .................................................................................................................. 113
  4.3.1.2 Procedure .............................................................................................................. 113
4.3.2 Results ......................................................................................................................... 113
4.4 General Discussion ......................................................................................................... 113

5. Nonsymbolic, approximate multiplication in monkeys (Jones, S. M. & Brannon, E. M.,
in preparation) ................................................................................................................... 118
  5.1 Experiment 1 .................................................................................................................... 127
    5.1.1 Methods .................................................................................................................... 127
      5.1.1.1 Subjects .............................................................................................................. 127
      5.1.1.2 Stimuli ............................................................................................................... 128
      5.1.1.3 Procedure ......................................................................................................... 130
      5.1.1.4 Training .......................................................................................................... 130
      5.1.1.5 Test Phase 1 .................................................................................................... 132
      5.1.1.6 Test Phase 2 .................................................................................................... 132
    5.1.2 Results .................................................................................................................... 133
      5.1.2.1 Trials to criteria .................................................................................................. 133
      5.1.2.2 Feinstein accuracy during test phases ............................................................. 134
      5.1.2.3 Testing alternative hypotheses ......................................................................... 136
  5.2 Discussion ..................................................................................................................... 139
6. Summary and Conclusions ........................................................................................................ 144

6.1 Chapter Summaries ............................................................................................................. 146

6.2 Discussion ........................................................................................................................... 152

6.2.1 Context .......................................................................................................................... 153

6.2.2 Cuing the two systems .................................................................................................... 158

6.2.3 Evolution of the ANS ...................................................................................................... 165

6.3 Conclusion .......................................................................................................................... 167

Appendix A ................................................................................................................................. 169

Appendix B ................................................................................................................................. 171

References .................................................................................................................................. 179

Biography .................................................................................................................................... 194
List of Tables

Table 1: Model fits for Chapter 2, Experiments 1A, 1B, and 2. Differences from maximum log likelihood (best fit) for each of four models. ................................................................. 54

Table 2: The object file system, the ANS, and the two system theory of numerical discrimination offer distinct predictions for performance on the conditions from Chapter 4, Experiment 1.................................................................................................................. 105

Table 3: Ecological differences in three lemur species which might predict differences in numerical acuity at a species level........................................................................................................ 147

Table 4: Chapter 4, Experiment 1 included a wide range of prosimian species, equally distributed between the numerical conditions. ................................................................. 169
List of Figures

Figure 1: Example stimuli in ordinal task. ................................................................. 46

Figure 2: Accuracy as a function of numerical ratio from Experiment 1. Unlike the results of Merritt, MacLean, Crawford & Brannon (2011), we found no prominent difference in accuracy between lemurs and monkeys. ......................................................... 51

Figure 3: Weber fractions ($w_{Pica}$) for the three lemur species and rhesus macaques from Experiment 1A, adult humans from Experiment 1B and the three extensively trained rhesus macaques from Experiment 2. Black circles represent medians of posterior distributions for individual weber fractions. Boxplots show quantiles of posterior distributions for species specific weber fractions in the Pica model. Distance between whiskers represents the 95% credible interval for the species-specific w......................... 56

Figure 4: All three subjects in Experiment 3 showed a decrease in weber fraction ($w_{Pica}$) between Time 1 and Time 2. ............................................................................................................ 68

Figure 5: (A) Example of an ascending trial in small range, (B) example of a descending trial in small range, (C) example of an ascending trial in large range, (D) example of a descending trial in large range. Note that the background screen color was red for all ascending trials and blue for all descending trials. ......................................................... 82

Figure 6: (A) Mean RTs for Mikulski in small range, (B) Mean RTs for Mikulski in large range, (C) Mean RTs for Schroeder in small range, (D) Mean RTs for Schroeder in large range. Both monkeys showed a semantic congruity effect in both ranges............... 86

Figure 7: Mean RTs averaged across both monkeys on the pair 6 vs. 12 for all correct trials in the small and large ranges. When the pair 6 vs.12 was the largest set of values in the small range, monkeys were, on average, faster to choose the larger than to choose the smaller of the two arrays. In contrast, when the pair 6 vs. 12 was the smallest set of values in the large range, monkeys were, on average, faster to choose the smaller than to choose the larger of the same two arrays. ................................................................. 88

Figure 8: Median RTs for descending trials for the pair 6 vs. 12 subtracted from median RTs for ascending trials for the pair 6 vs.12. Data reflects the average of the median RTs for correct trials for the 5 sessions of each block averaged for the two monkeys. Positive values indicate that the monkeys were faster on descending trials compared to ascending trials for the pair 6 vs. 12, whereas negative values indicate the reverse. Monkeys showed a rapid contextual shift between blocks such that in blocks of small
range sessions, monkeys were faster on descending trials, and in blocks of large range sessions, monkeys were faster on ascending trials for the 6 vs. 12 pair.

Figure 9: Subjects made a choice by approaching one of the two buckets.

Figure 10: Results support the predictions of the ANS. Subjects successfully chose the larger in a small vs. large discrimination for conditions in which values differed by a 1:3 ratio, but not a 1:2 ratio.

Figure 11: Subjects chose the larger of the 1 vs. 2 (nonsignificant trend, $p = 0.06$) and 1 vs. 3 ($p < 0.01$) conditions in Experiment 1, but failed to choose the larger of 2 vs. 3 in Experiment 2, implicating the ANS rather than the object file system.

Figure 12: Example of 2+2 addition movie clip. Arrows indicate the direction of movement of the squares.

Figure 13: Example of 2x2 multiplication movie clip. Arrows indicate the direction of movement of the squares.

Figure 14: Feinstein was tested, without reward, on her ability to perform arithmetic calculations on probe trials in which familiar numerical arrays (1 or 2 dots) were presented in a novel number of green squares (3). Her performance indicated that she extended an arithmetic rule to novel stimuli for both addition and multiplication trials.

Figure 15: Chapter 5, Experiment 1. Addition Stage 1: 2+2

Figure 16: Chapter 5, Experiment 1. Addition Stage 1: 2+2+2+2

Figure 17: Chapter 5, Experiment 1. Addition Stage 2: 2+2

Figure 18: Chapter 5, Experiment 1. Addition Stage 2: 2+2+2+2

Figure 19: Chapter 5, Experiment 1. Multiplication Stages 3 through 5: 2x2

Figure 20: Chapter 5, Experiment 1. Multiplication Stages 3 through 5: 2x4

Figure 21: Chapter 5, Experiment 1. Addition Stage 5: 2+2

Figure 22: Chapter 5, Experiment 1. Addition Stage 5: 2+2+2+2
1. Introduction

1.1 Number: A case study for examining abstract thought without language

The nonsymbolic representation of number is an ideal case study for probing the evolutionary development of nonlinguistic thought. Number is invariant across all possible discrete elements; three peanuts, three elephants, and three dissertations are equally good exemplars of threeness. Thus, numerosity is an abstract property of a set. In studying the approximate numerical abilities ubiquitous across many species, we have the potential to learn about the evolutionary origins of abstract, non-linguistic thought.

Approximate numerical representations are abstract in that they are characterized by independence from the physical and perceptual qualities of the individual items that comprise a set. Such representations appear to be normalized over continuous properties of the set such as size, area, perimeter, and density. In addition, featural properties such as color and shape are irrelevant for the representation of number. Furthermore, nonhuman primates have been shown to represent, compare and combine the numerosity of elements in a set even when presented in multiple sensory modalities. For example, Jordan, MacLean and Brannon (2008) trained two rhesus monkeys to choose an array of 1-9 squares that numerically matched a sequence of visual or auditory stimuli. They matched numerosity across modalities (a sequence of sounds to a visual array) as accurately as they matched within modality (a sequence of
visually presented shapes to a simultaneous visual array). Additionally, when the
sample sequence consisted of both visual and auditory elements, monkeys were able to
select a visual array that matched the numerosity of the intermixed sequence.
Performance was ratio-dependent, indicating that approximate numerical
representations are independent of modality.

Although physical and perceptual features of the stimuli do not appear to be
retained as information in the approximate number system (ANS), nonhuman primates
can selectively attend to individual elements of one color or shape to form ANS
representations. For example, Beran (2008a) demonstrated that rhesus macaque
monkeys were able to select the larger of two sets of black dots in a computerized
display, while ignoring the spatially intermixed red dots. Rhesus macaques performed
above chance when the largest subset of black dots was embedded in the smaller
intermixed array. Thus, the rhesus macaques in this study were able to use identity
information (color) to determine which subset to represent numerically. On the other
hand, Cantlon and Brannon (2006a) reported no cost to reaction time or accuracy when
comparing heterogeneous arrays rather than homogenous arrays. Thus, while a subset
may be selected for numerical representation using physical features or identity
information, those features do not appear to be encoded by the ANS. Note that while a
numerical representation may be stored with identity information when it passed into
memory, the magnitude representation itself is abstract.
A vast and diverse literature suggests that humans and nonhuman primates share an evolutionarily ancient system of nonverbal number representation. Recent work suggests that this system may also be common to taxa as diverse as fish (e.g. Agrillo, Dadda, & Bisazza, 2007; Agrillo, Dadda, Serena, & Bisazza, 2008; Agrillo, Piffer, Bisazza, & Butterworth, 2012) and birds (e.g. Emmerton, 1998; Scarf, Hayne, & Colombo, 2011). This research has largely focused on providing existence proofs of numerical competence or comparing individual species with humans. In the following chapter, I will address the contextual and quantitative limits of the approximate number system in nonhuman primates, with a particular focus on the implications of context and species differences and similarities on the evolution of numerical cognition.

1.2 A shared system of approximate numerical representation

Humans use numerical information to perform a wide variety of complex tasks, from balancing a checkbook to designing a skyscraper. However, this use of exact number relies on symbolic representations such as count words (one, two, three, etc.) or Arabic numerals (1, 2, 3, etc.) as discrete representations of the values for which they stand. Non-human animals cannot represent exact numerical values and I argue that this is because they lack symbolic language. In contrast to the language-dependent precise system for representing number that is unique to humans, human and non-human animals share a system for representing numerical values in an approximate format.
1.2.1 Ratio Dependence

Approximate number discrimination is governed by Weber’s law: accuracy decreases and reaction time increases for numerical comparisons as the ratio between the two values (min/max) approaches 1. For adult humans, Weber’s law holds across numerical formats such as words, Arabic numerals, tones, and dot arrays (Dehaene & Akhavein, 1995; Dehaene, Izard, Spelke, & Pica, 2008; Pica, Lemer, Izard, & Dehaene, 2004; Whalen, Gallistel, & Gelman, 1999). Weber’s Law states that, for perceptual continua such as loudness or duration, $\text{DI}/I = C$, where DI is the increase in intensity to a stimulus of intensity I that is required to produce a detectable change and C is a constant. Thus, the change in stimulus intensity needed for a change to be detectable is a constant proportion of the original stimulus intensity rather than a constant amount. For example, if an increment of 2cm is needed to detect a change in a 20cm line, then an increment of 4cm would be needed to detect a change in a 40cm line.

Weber’s Law applies to numerosity discrimination in a similar way in which it characterizes perceptual continua. Moyer and Landauer (1967) first showed ratio dependent performance in reaction time (RT) for adults’ comparisons of Arabic numerals. In a digit-pair discrimination task, participants were asked to use a left or right switch to indicate whether the digit presented on the left or the right of a display was numerically larger. The response functions obtained in this study indicated that, like in classic psychophysical functions, RT was modulated by the ratio between two
digits: RT decreased with increasing numerical distance between two values, and if
distance is held constant, RT increases with numerical magnitude. These results
provided the first strong evidence that numerical comparisons in adult humans are
subject to Weber’s law.

These ratio effects in human number discrimination have been replicated in
many different languages and across multiple formats. In fact, such approximate
representations of number are language-independent as evident from studies on
nonsymbolic numerical discriminations by members of two indigenous Amazonian
cultures with languages that contain few number words. The Piraha have words for
‘one’ and ‘two’, but use a single term for ‘many’ to refer to all larger values. When asked
to produce sets of items that numerically matched an example set, Piraha participants
successfully produced sets of one or two items, but showed approximate performance
for larger set sizes. Similarly, the Munduruku language contains only number words for
small values (1–5). When Munduruku participants compared the numerosity of large
numbers of dots (20–80), their performance was similar to that of French-speaking
control participants. Accuracy was modulated by the ratio between the two values.
Thus, adult humans possess language-independent magnitude representations of
number that are dependent on the ratio between the numerical sets.

Studying nonverbal discrimination or estimation of numerosities in adults whose
language allows them to symbolically represent number presents a challenge given that
adults are likely to verbally count. Whalen and colleagues (1999) were able to prevent participants from counting by presenting stimuli and requiring responses that were too rapid for verbal counting. Cordes, Gelman, Gallistel and Whalen (2001) used a different method of controlling counting: participants were required to speak out loud while pressing a key at faster than five presses per second. This prevents subvocal counting due to speed and an alternate use of language. Cordes and colleagues (2001) modified a task from the animal literature (Platt & Johnson, 1971), such that participants were asked to make between 7 and 25 key presses as fast as they could. They demonstrated that even when verbal counting was eliminated as a possibility, adult humans’ nonverbal number representations showed scalar variability.

Brannon and colleagues (e.g., Brannon & Terrace, 1998, 2000; Cantlon & Brannon, 2005) have suggested that the evidence for approximate number discrimination in monkeys may indicate a shared system of numerical representation in humans and nonhuman animals. Recent studies have directly compared the numerical abilities of humans with that of other species, providing strong evidence in favor of this claim. Here I will review a few such studies, comparing the numerical abilities of humans with other primate species.

Beran, Johnson-Pynn and Ready (2008) and Jordan and Brannon (2006a,b) both used a bisection task to compare approximate numerical performance in children and rhesus monkeys. Children and rhesus monkeys were trained to respond differently to
two anchor values, and were tested with intermediate values. Beran, Johnson-Pynn and Ready (2008) trained participants to associate anchor values with symbols, while Jordan and Brannon (2006a,b) used arrays as anchor values rather than symbols. The results suggest that the ANS underlies numerical discrimination in both species for both symbolic and nonsymbolic anchors.

Cantlon and Brannon (2006b, 2007a) systematically compared approximate numerical abilities in rhesus monkeys and human adults using identical tasks. Cantlon & Brannon (2006b) reported that monkeys successfully learned an ordinal rule for numerical arrays. A direct comparison of humans and monkeys on this task showed qualitative and quantitative similarities in performance indicative of a shared system of nonverbal number representation. Cantlon and Brannon (2007a) also used a match-to-sample task to compare nonsymbolic addition in rhesus macaques and adult humans. Two sample arrays of dots were presented sequentially on a computer screen. After a brief delay, participants had to choose which of two numerical arrays matched the sum of the two samples. Performance by both species was modulated by the ratio between the choice arrays, which demonstrates humans and monkeys are capable of summing magnitude representations of sequentially presented sets.

Brannon and Terrace (1998) trained two rhesus monkeys to order visual arrays of 1-4 items in ascending order. The monkeys learned to order arrays controlled for size, shape and color to receive a food reward. Both monkeys demonstrated spontaneous
ordering for pairs of the numerosities 5-9, although they had never been reinforced for ordering those values. As well as demonstrating that monkeys are capable of learning an ordinal rule, these results also showed a distance effect similar to that reported by Moyer and Landauer (1967). Brannon and colleagues (Brannon & Terrace 2000; Brannon, Cantlon, & Terrace, 2006) extended these findings to descending ordering.

In these early ordinal studies by Brannon and colleagues, each monkey was trained on a single ordinal rule. Cantlon and Brannon (2005) modified the ordinal task for use in a within-subjects design. Rhesus macaques were trained to order a pair of numerosities in ascending or descending order, depending on the color of the display background. When the background was red, monkeys were rewarded for selecting the smaller array first. When it was blue, they were rewarded for selecting the larger array first. Density, cumulative surface area, and cumulative perimeter had less, if any, effect on responding than numerosity. After training, the monkeys were tested on all pairs 1-9 in both ascending and descending order. Results show a clear pattern of ratio dependence.

Many non-primate species from a wide variety of taxa as distantly related as fish (e.g. Agrillo, et al., 2008) and rats (e.g. Meck & Church, 1983) show ratio-dependent numerical discriminations. However, although many different animal species have been tested in a wide variety of tasks, rarely have multiple species been tested with the same task and stimuli to allow for a quantitative comparison. A recent study suggests that
there are no qualitative differences in the numerical discrimination abilities of rhesus macaques and pigeons (Scarf, et al., 2011). In that study pigeons successfully transferred an ordinal numerical rule from the values 1-4 to the values 5-9 much like rhesus monkeys (Brannon & Terrace, 1998). Pigeons showed ratio dependent performance that looked quite similar across these two species separated by over 200 million years of evolutionary history.

Ratio dependent performance in numerical discrimination in accordance with Weber's law, as is evidenced by the wealth of data on human and nonhuman nonsymbolic numerical discrimination, is the hallmark signature of the ANS. According to Weber's law, the increase in intensity required to detect a change to a stimulus is a function of the magnitude of the initial intensity. Specifically, the numerosity of a set of n dots is represented internally by a Gaussian random variable X with mean q(n) and with standard deviation w(n). Thus, we are able to specify the precision of internal quantity representations by calculating the internal weber fraction, w. However, it is also important to note that weber fractions are not hard and fast invariants. The absolute value will vary based on the model used to estimate w and may also vary based on a multitude of different factors such as the duration of stimulus presentation or the range of numerical values used. These factors mean that it is essential to use the same task parameters when attempting to make species comparisons.
Several models have been developed to estimate weber fraction for numerical discrimination. One model, presented by Pica, Lemer, Izard and Dehaene (2004), assumes that the numerosity of a set of n dots is represented internally by a Gaussian random variable. This model includes one free parameter: the internal weber fraction w, and assumes a linear internal number line with scalar variability. An alternative model, presented in Piazza, Izard, Pinel, Le Bihan and Dehaene (2004), assumes that subjects have an exact representation for some baseline value; w is calculated using the overlap of distribution functions of internal representations for each test value with the exact baseline representation. The Piazza model also includes a single free parameter, w, but assumes a logarithmic internal number line with fixed Gaussian variability (although, note that a linear number line with scalar variability predicts virtually identical results).

1.2.2 The Semantic Congruity Effect — another hallmark of the approximate number system

In addition to ratio dependence, another characteristic of approximate numerical comparisons is the semantic congruity effect. When asked to compare the size of two large items, adult humans are quicker to respond to the question “Which is larger?” than “Which is smaller?” Conversely, when asked to compare two small items, adults are quicker to respond to the question “Which is smaller?” than “Which is larger?” The semantic congruity effect has been demonstrated with a variety of stimuli, including physical sizes, distances, line length, brightness, surface area (Audley & Wallis, 1964; Čech & Shoben, 1985; Čech, Shoben, & Love, 1990; Holyoak & Mah, 1982; Moyer &
Bayer, 1976; Petrusic, Baranski, & Kennedy, 1998). This effect has also been found with numerical comparisons of Arabic numerals and dot arrays (Banks, Fujii, & Kayra-Stuart, 1976; Holyoak, 1978). When evaluating Arabic numerals, adults indicate that 2 is smaller than 3 more quickly than they indicate that 3 is larger than 2. However, when presented with two relatively large values such as 8 and 9, adults are faster when asked to indicate the larger compared to the smaller value (Banks, et al., 1976).

Cantlon and Brannon (2005) demonstrated a numerical semantic congruity effect with a nonverbal species: the macaque monkey (Macaca mulatta). Monkeys were trained on a touch screen computer to choose between two arrays of dots. A color cue indicated whether they would be rewarded for selecting the array with the larger (blue background) or smaller (red background) number of dots. Monkeys showed a semantic congruity effect similar to that reported for humans: with a red background, they were faster to respond that 2 was smaller than 3 than they were to respond that 3 was larger than 2. In contrast, when the values were relatively large, monkeys were faster to indicate the larger of the two values than they were to indicate the smaller of the two values with a blue background.

1.3 **Approximate number system and food choice tasks**

A large literature exists on the ability of nonhuman animals to use approximate representations. For example, the ability to select the larger of two food quantities has been demonstrated in several species of primates besides humans (e.g. Addessi,
Animals are motivated to choose the larger amount of food. In most cases (when food items are a consistent size), number of food items covaries with amount of food. However, when using food items of various sizes, animals are motivated to select the larger amount even when it is the smallest number of pieces (Stevens, Wood, & Hauser, 2007; Beran, Evans, & Harris, 2008).

This food choice task has been used to provide a direct comparison of numerical discrimination in several ape species. Hanus and Call (2007) investigated the ability to select the larger of two quantities in bonobos, chimpanzees, gorillas, and orangutans. Participants watched as food pellets were either sequentially added to opaque containers, simultaneously presented and visible throughout selection, or presented successively (one set is displayed, then covered before the other set is displayed). In all presentation types, apes showed ratio-dependent performance such that they were increasingly successful at selecting the larger number of food items as ratio (small number/larger number) decreased. Anderson, Stoinski, Bloomsmith and Maple (2007) extended this finding to include a summation condition, in which orangutans were choosing between two pairs of containers, such that they had to sum the amount in each pair to determine the total number of grapes in each set. Performance on this summation condition was also ratio-dependent.
Monkeys have also been shown to select the larger of two sets of food items using approximate numerical representations. vanMarle, Aw, McCrink and Santos (2006) showed ratio dependence in numerical comparisons with capuchin monkeys. Although the pairs were all within the small number range (1-4), participants’ selections showed ratio-dependence. Beran, Evans, Leighty, Harris and Rice (2008) added to this finding by showing that capuchin monkeys can select the larger of two sets when the sets contain larger numbers (1-6 items).

The issue inherent in using food items as stimuli is that total amount of food must necessarily covary with the number of food items for subjects to be motivated to select the larger number. Carefully designed computerized tasks make it possible to adequately control other stimulus dimensions, which allow for a stronger argument that animals discriminate sets based on numerosity. Multiple studies have explored the abilities of nonhuman primates to order and match sets of computerized stimuli based on numerosity, while controlling for confounding variables.

1.4 **Do animals use number in the wild?**

Optimal foraging theory predicts that when faced with alternative foraging options, animals will choose the one that provides the greatest rate of return. During foraging, animals must decide at what point they should stop extracting resources from the current location and move on to search for a new one. Optimal foraging theory makes predictions about how long to stay: until the rate of return at a location drops
below the baseline rate in the environment (Stephens & Krebs, 1986). Research on foraging behavior indicates that animals spontaneously determine rates of return. This requires that animals approximate the amount of food obtained, the time spent obtaining it, or some combination of amount and time.

Research with species as phylogenetically distant as insects and birds has also provided evidence that animals spontaneously utilize numerical information in foraging (e.g. Bar-Shai, Keasar, & Shmida, 2011; Farnsworth & Smolinski, 2006; Hunt, Low, & Burns, 2008; Tsutsumi, Ushitani, & Fujita, 2011). For example, Bar-Shai, et al. (2011) made use of the fact that numerical regularity in spatial distribution occurs naturally in some flowers. Bees that forage for nectar on such flowers can increase efficiency by avoiding return visits to empty nectaries. Both observational studies and laboratory tests indicated that bumblebees adapted their behavior to the number of food sources. Farnsworth and Smolinski (2006) used a novel foraging task to assess numerical competence in mockingbirds. They designed double-ended feeders that were blocked by different numbers of bamboo sticks at each end. Subjects could access food by removing all sticks from either end. If subjects were able to utilize numerical information to maximize foraging efficiency, they should prefer to access the end with fewer obstacles. Results indicated that mockingbirds preferred the smaller number of obstacles in 1 vs. 6 and a 2 vs. 5 conditions, but chose at chance on a 3 vs. 4 comparison,
suggesting that they utilized numerical information in foraging decisions using approximate representations.

Further evidence for the use of number by animals in the wild comes from naturalistic studies, which rely on the preferences and behaviors of untrained subjects. Shoaling is such a behavior that has been used to test the numerical discrimination abilities of fish. When in a new environment, female mosquito fish will join a shoal, or collection, of other female fish for protection. Agrillo and Colleagues (2007, 2008) made use of this spontaneous behavior to study numerical discrimination. This set of studies which consisted of two conditions: adjacent numerical comparisons between 1 and 8 (e.g., 2 vs. 3, 3 vs. 4), and comparisons of large numbers (>3) with more discriminable ratios. Results showed evidence of ratio-dependent performance with large values, such that they reliably joined the numerically larger of two shoals when shoal size differed by a 1:2 ratio (e.g. 8 vs. 16), but failed to do so when shoal size differed by a more difficult 2:3 ratio (e.g. 8 vs. 12). In addition, Agrillo and colleagues argued that the fish used a separate system to discriminate small sets. This alternate system of numerical discrimination will be discussed in more detail to follow.

Another spontaneous behavior used to study number discrimination in naturalistic settings is assessment of the number of animals in an unfamiliar group and in one’s own group during territorial disputes. Wilson, Hauser, and Wrangham (2001) relied on the spontaneous behavior of male chimpanzees to escalate aggression towards
an unfamiliar male depending on numerical assessment, with lethal attacks occurring when numerical advantage reduces the costs of attacking. In a series of playback experiments they tested whether the response to the playback of the call of a single out-group male depended on the number of males in the listening party. They found that groups of three or more males consistently vocalized and approached the speaker together, while those with fewer adult males usually stayed silent and approached the speaker less often. McComb, Packer, and Pusey (1994) reported a similar finding with female lions. Experiments simulating the intrusion of unfamiliar females found that female lions were more likely to approach a single intruder’s roar if they outnumbered the intruder, and were more likely to approach a single roar than three strangers’ roars.

There is also limited evidence that animals are able to represent numerical information about a subset of items in a mixed visual array. Lyon (2003) indicated that discriminating number may be relevant for bird species vulnerable to brood parasitism. Coots that recognize parasitic eggs mixed with their own base clutch-size decisions on only their own eggs, while coots that mistake parasitic eggs as their own include these in their numerical assessment. The ability of female coots to count only their own eggs in a mixture of eggs provides an example of a spontaneous ability to represent the numerosity of a subset of objects in a mixed array.
1.5 Comparative Cognition and the case for studying prosimians

Research on numerical cognition in nonhuman primates has largely focused on chimpanzees and monkeys (e.g., Addessi, et al., 2008; Anderson, et al., 2007; Beran, 2008a,b,c; Beran, et al., 2008; Brannon & Terrace 1998, 2000; Brannon, et al., 2006; Cantlon & Brannon, 2005, 2006a,b, 2007a,b; Hanus & Call, 2007; Jordan, et al., 2008; vanMarle, et al., 2006). Much less research has been conducted on the cognitive abilities of prosimians primates. Including the prosimian suborder of primates (lemurs, lorises and galagos) in comparative research is likely to be important in attempting to identify cognitive profiles of the primate ancestral state: prosimians have been hypothesized to be similar to the last common primate ancestor (Tattersall, 1982; Yoder, 2007). If prosimian primates demonstrate cognitive traits that are common among other primates, it is likely that these traits were present in the last common ancestor.

The limited data seem to indicate that lemurs differ from apes and monkeys both quantitatively and qualitatively on measures of social intelligence (for a review, see Fichtel & Kappeler, 2010). Alternatively, recent studies have found qualitative similarities between lemurs and monkeys in non-social cognitive tasks. Lemurs performed more like macaque monkeys than pigeons in a sequence learning task (Merritt, McLean, Jaffe & Brannon, 2007). Additionally lemurs share with monkeys a capacity for cognitive abilities such as Piagetian object displacement (Deppe, Wright & Szelistowski, 2009), transitive reasoning (MacLean, Merritt & Brannon, 2008), and spatial
memory (Erhart & Overdorff, 2008; Lührs, Dammhahn, Kappeler, & Fichtel, 2009). The few studies that examined numerical abilities in lemurs indicate a capacity for approximate number discrimination comparable to that of other nonhuman primates.

Santos, Barnes, and Mahajan (2005) investigated the spontaneous representation of number in lemurs with a violation of expectancy paradigm. Lemurs watched as two lemons were hidden behind a screen; the screen was then lowered to reveal either a numerically possible outcome (two lemons), or an impossible outcome (one lemon or three lemons). On average, lemurs looked longer at the impossible outcome than at the possible outcome, suggesting an ability to represent the number of lemons behind the screen. Lewis, Jaffe, and Brannon (2005) also found evidence that lemurs represent the numerosity of hidden objects. In this study, mongoose lemurs watched as grapes were hidden in a bucket that was filled with shredded paper. A subset of these grapes was hidden in the container’s false bottom. After lemurs were allowed to retrieve and consume the available grapes, the additional time they spent searching in the bucket was measured. If lemurs expected more grapes to remain in the bucket, they should have continued to manually search after retrieving the available ones. Results indicated that search time was dependent on the number of grapes that should have been in the bucket and reflected ratio dependent discrimination.

Merritt, MacLean, Crawford, and Brannon (2011) extended this finding of numerical representation to show that lemurs are able to learn a numerical rule and
extend that rule to novel stimuli. Two ring-tailed lemurs were trained to select the numerically smaller of two arrays of dots (values 1 to 4), using trial-unique stimuli and surface area controls. After successfully learning to order pairs of the values 1-4, the lemurs were able to extrapolate this ordinal rule to novel numerical values (arrays of 5 to 9 dots) replicating the basic pattern of results found by Brannon and Terrace in macaque monkeys (Brannon & Terrace, 1998, 2000). While lemur performance in this ordinal numerical comparison task was qualitatively similar to that of monkeys, accuracy was generally lower for the ring-tailed lemurs compared to the macaques.

1.6 Two systems view

As mentioned previously, not all spontaneous numerical discriminations show the hallmark of ratio dependence. Instead, there is evidence for a second numerical system that allows for exact numerical representation of small sets. Preverbal human infants and nonhuman animals show these two contrasting patterns of results for numerical discrimination. Successful discrimination of small values only has been taken as evidence for the object file system, which allows the accurate representation of a small number (less than 3 or 4) of objects.

1.6.1 Object File System

Not all spontaneous numerical discriminations show the hallmark of ratio dependence. Instead, there is evidence for a second numerical system that allows for exact numerical representation of small sets. Preverbal human infants and nonhuman
animals show these two contrasting patterns of results for numerical discrimination. Successful discrimination of small values only has been taken as evidence for the object file system, which allows the accurate representation of a small number (less than 3 or 4) of objects.

The object file system was originally proposed as a model of visual perception in adults. The use of object files allows us to track of objects despite changing physical features and without requiring us to semantically identify or label the individuals (Kahneman, Treisman, & Gibbs, 1992). Adult humans form these object files as a mechanism for tracking visual objects in working memory according primarily to their spatiotemporal properties. The object file system represents individuated objects, with the number of open object files providing an implicit way to represent the numerosity of a set. Object files are not explicitly numerical representations, but instead represent individual objects in attention. Each object file “sticks” to a unique object as it moves about the visual scene, and may contain identity or featural information. The object file system represents individuated objects, with the number of open object files providing an implicit way to represent the numerosity of a set. As only 3 or 4 object files can be maintained simultaneously, the ability of this system to provide a means of representing numerosity is limited to small numbers. The ANS represents the cardinality of a set of objects as a single mental magnitude. The ability to discriminate between two numerosities in the ANS is subject to ratio-dependence, in accordance with Weber’s
Law. These different kinds of representation appear to be incompatible in a single discrimination (vanMarle, 2012).

1.6.2 Developmental literature

A food choice task used has repeatedly produced a set-size limitation in the numerical discriminations of infants (Feigenson & Carey, 2005; Feigenson, Carey, & Hauser, 2002). In this paradigm, infants watch as food items are dropped into two containers and are then allowed to crawl to one of the containers and consume its contents. Feigenson and colleagues (2002) demonstrated that 10 to 12 month old infants reliably crawled to the container with the greater number of food items when both containers had 3 or fewer food items. Thus, 10 and 12 month old infants succeeded at choosing the larger in a 1 vs. 2 and a 2 vs. 3 condition, but selected at chance in a 2 vs. 4 or a 3 vs. 6 item discrimination. Controls for overall duration, number of reaches into the bucket, and satiation caused no change in performance. In a separate experiment, they demonstrated that infants performed at chance in a 1 vs. 4 item discrimination, but successfully chose four over zero crackers, indicating that infants were capable of representing the existence of food items, even if they couldn’t represent the actual number (Feigenson & Carey, 2005).

A manual search task also produced this set size limit in the numerical discriminations of 10 to 14 month old infants (Feigenson and Carey, 2003, 2005). When 3 toys were placed in a box and the experimenter retrieved 2, infants spent more time
searching for the missing toy than when 2 toys were placed in the box and the experimenter retrieved both. This difference in search time indicates that infants represented the number of toys in the box and the number of toys already removed, and were able to differentiate between those values. In contrast, when 4 toys were placed in the box and 2 were retrieved, infants did not search longer than when all items had been removed from the box, suggesting that they could not discriminate 2 from 4 items. Together, these findings indicate that under some circumstances, once the number of objects in a set exceeds the object file limit of 3, infants can no longer discriminate between sets on the basis of numerosity (even in a singular/plural distinction.)

1.6.3 Evidence from animal literature

Limited evidence has suggested a similar object file system for nonhuman animals. Hauser, Carey and Hauser (2000) found a set size limit in the spontaneous numerical discriminations of rhesus monkeys using a single-trial food choice task similar to the task used by Feigenson and colleagues (2002). Free ranging, untrained rhesus macaques were shown apple slices being placed into two boxes. Monkeys were allowed to approach and consume the apple slices in one box. Monkeys consistently chose the greater number of apple slices as long as the contents of each box did not exceed the set size limit of 4. However, on trials where one box did exceed that limit (with the exception of one pair: 3 vs. 5), the monkeys approached the boxes at random, showing no preference for greater number of food items. This was true even with
favorable ratios: 4 vs. 8 and 3 vs. 8. Wood, Hauser, Glynn and Barner (2008) also found evidence for a set size limit in the numerical discrimination abilities of rhesus macaques, this time with nonsolid food portions. Monkeys spontaneously approached the box containing the larger number of scoops of carrot pieces, poured from a cup, for 1 vs. 2, 2 vs. 3, and 3 vs. 4 scoops, but not 4 vs. 5 or 3 vs. 6.

Uller and Lewis (2009) performed an experiment similar to the one reported by Hauser and colleagues (2000) demonstrating quantity discrimination in horses. Horses watched as apples were placed in each of two buckets. Horses chose the larger of 1 vs. 2 and 2 vs. 3 apples, but failed to systematically choose the larger number of apples in a comparison of 4 vs. 6 apples. Red-backed salamanders showed a similar pattern in a food choice task: they chose the larger set when presented with 2 vs. 3 fruit flies, but not 4 vs. 6 fruit flies despite the ratio equivalence (Uller, Jaeger, Guidry, & Martin, 2003).

Using a training paradigm, Rugani, Regolin, & Vallortigara (2008) found evidence for the object-file system in newly-hatched chickens. Young chicks were trained to peck at arrays of dots depending on their numerosity. In a between subjects design, chicks were successfully trained to discriminate arrays of 1 vs. 2 and 2 vs. 3 dots, but failed to learn to discriminate arrays of 3 vs. 4, 4 vs. 5, or 4 vs. 6 dots. Controls for area, contour length, and density indicate that the chicks discriminated the stimuli based on number and not a continuous variable.
1.6.4 Factors that may influence the recruitment of the ANS vs. object files

vanMarle and colleagues (2006) argued that evidence for the ANS was often found with animals which had undergone extensive training, which may cause animals to represent quantity differently than they would do spontaneously. While object files do appear to be more likely to be recruited in single-trial discrimination tasks (e.g., Hauser, et al., 2000; Uller & Lewis, 2009), single-trial tasks are not sufficient to always recruit the object file system (Agrillo, et al., 2007, 2008; Flombaum, Junge, & Hauser, 2005). Conversely, studies involving training elicit approximate numerical representations (e.g., Beran, 2007, 2008a,c; Cantlon & Brannon, 2006b, 2007a). Thus, numerical training may eliminate any propensity to elicit the object-file system given that there are no published studies in which numerically trained animals show a set size limitation.

Large numerosities can only be handled by the ANS, since they exceed the set size limit of the object file system. It would make intuitive sense that the presentation of large quantities would automatically recruit the ANS. However, there is evidence that, even in circumstances when the ANS would allow success in a task that animals and infants may not be able to inhibit object file representations. Several studies have found that animals and infants demonstrate the set size limitation even with pairs that should be highly discriminable in the ANS (e.g. Feigenson & Carey, 2003, 2005; Hauser, et al., 2000).
Hyde (2011) proposed that conditions that allow attentional selection of individual elements cue the object file system, while conditions that inhibit the ability to attend to individual elements will result in approximate numerical representations. Barner, Wood, Hauser and Carey (2008) and Hyde and Wood (2011) each proposed ways to experimentally manipulate which system is recruited. Hyde and Wood (2011) argued that spatial attention to individual elements is required for representation by the object file system but not the ANS. They found that when the viewing angle of an array was within the spatial resolution of attention, participants detected a change in numerosity for small sets (<4) with high accuracy and no ratio-dependence. However, when the viewing angle was reduced to beyond the spatial resolution of attention, accuracy was modulated by the ratio of numerical change. Barner, Wood, Hauser and Carey (2008) argued that 1. If sets contain a small number of items and are presented only once, without training or familiarization, the ANS may not be spontaneously activated. 2. If sets are large, if individual elements do not move independently (e.g., are static or move as a unit), or if sets are presented repeatedly through training or extensive familiarization, the object file system may not be spontaneously activated. Using numerical arrays designed to elicit neither representation, Barner and colleagues argued that, when neither system is recruited, rhesus macaques failed to show performance consistent with either the ANS or the object file system.
Thus, while some candidates have been proposed, it is unclear exactly what aspects of a task determine which system of numerical discrimination is recruited. What is clear is that these two systems do not reflect a simple discrepancy between small and large values, as there is evidence for approximate representation in the small range (e.g. Brannon & Terrace, 1998; Merritt, et al., 2011; vanMarle, et al., 2006). Nor do these systems characterize a simple discrepancy between trained and spontaneous performance, as there is evidence for ratio dependence in spontaneous discriminations (e.g. Agrillo, et al., 2007, 2008; Flombaum, et al., 2005). Instead, the relationship between these systems appears to be more complicated, cued by task demands and experience of the subjects.

1.7 Last resort hypothesis

In order to disambiguate numerical discrimination from discrimination based on continuous variables, it is necessary to utilize careful stimulus controls. When physical objects in a set are equal in size, other variables such as cumulative surface area and cumulative contour length covary with numerosity. Likewise, when auditory events in a sequential set have equal duration, overall duration covaries with numerosity. To determine whether subjects are discriminating between two sets on the basis of number, or whether a continuous variable is responsible for response patterns, these confounding variables must be controlled for.
One early perspective was that animals could be trained to make numerical discriminations, but that the use of numerical information was both difficult and unnatural. Davis and Memmott (1982) argued that numerical competence in animals developed slowly in response to task demands, but was not a naturally occurring ability. According to this perspective, nonhuman animals could attend to numerosity only as a “last resort”, when all alternate variables (such as size, color, shape or duration) were unavailable or uninformative. Thus, animals were believed to be capable of representing the numerosity of a set, but this capacity emerged only through training or as a “last resort” in the absence of other salient properties.

An alternative perspective is that animals routinely and automatically encode number. In support of this perspective, Meck and Church (1983) showed that when animals were trained to discriminate stimuli that differed in both the duration and number of events, they encoded both time and number even though it would have been sufficient to encode either one. Rats were trained to press one lever after a two-cycle white noise with a 2-second duration and to press a different lever after an eight-cycle white noise with an 8-second duration. A cycle was defined as .5-second sound-on and .5-second sound-off. After training, subjects were tested with stimuli that held one variable constant. They were presented with a 4-second stimulus with 2, 4, 6, or 8 cycles or 4 cycles that lasted for a total of 2, 4, 6, or 8 seconds. The rats encoded both duration and number, as evidenced by their bisection functions for each dimension. The rats
exhibited similar Weber fractions for duration and numerosity, indicating a similar sensitivity to changes in duration and numerosity. Thus, without being explicitly trained to do so, rats were found to encode number, despite the fact that encoding duration alone would have been sufficient.

Cantlon and Brannon (2007b) provided strong evidence in favor of the perspective that animals attend to numerical information automatically. They trained number-naive and number-experienced rhesus macaques to select a visual match to a sample stimulus to obtain a reward. For all reinforced trials, the sample and the rewarded match were equal in numerosity as well as a non-numerical dimension such as cumulative surface area. On non-reinforced probe trials, the choice stimuli each matched the sample on a different dimension, such that subjects could select a match based either on numerosity or on other continuous variables. All monkeys showed a bias to base their decision on number over cumulative surface area when the numerical ratio was favorable. Contrary to the perspective that number is a “last resort”, these studies provide evidence that animals attend to number regardless of training, even when other relevant dimensions are available.

1.8 Mental arithmetic in nonhumans

1.8.1 Ordering

Ordinality is perhaps the most basic form of mental arithmetic, and is intrinsic to the development of verbal counting (e.g. Davis & Pérusse, 1988; Gelman & Gallistel,
The ability to order numerical values requires the ability to appreciate that 2 is numerically more than 1 rather than simply discriminating that 2 is different from 1. According to Gallistel and Gelman (1992), this illustrates the difference between having number categories (the ability to represent the numerosity of a set independent of its other attributes) and having a number concept (the ability to manipulate sets through ordering, addition, subtraction, multiplication or division). Thus, having such a numerical concept allows one to reason arithmetically.

Many species have been shown to select the larger of two sets, but fewer studies have carefully studied the ability of nonhuman subjects to order numerical stimuli. Brannon and Terrace (1998, 2000) trained rhesus macaques to order visual arrays of 1 to 4 items based on their numerosity independent of other continuous variables. Monkeys were then tested on novel values 5-9 to demonstrate transfer of the ordinal rule. If they had memorized the pairwise responses to order the small values (a numerical category) this rule-learning would fail to transfer to novel numerosities. Instead, they found evidence that monkeys were able to transfer the ordinal rule, indicating a number concept. Merritt, et al. (2011) found similar results with ring-tailed lemurs, as did Scarf, et al. (2012) with pigeons, indicating that understanding ordinality might be ubiquitous across nonhuman species.

Gallistel and Gelman (2000) argue that the evidence that nonhuman animals order, add, and subtract numerosities indicate that, animals reason arithmetically, as the
axioms defining these operations define the system of arithmetic. In mathematical theory, the natural numbers are defined by the Peano axioms. These axioms can be summarized as such: 1. Zero is a number, 2. if n is a number, the successor of n is a number, 3. zero is not the successor of a number, 4. if two successors are equal, the numbers they succeed are themselves equal, and 5. if a set S of numbers contains zero and also the successor of every number in S, then every number is in S (Wolfram, 2002). The result of these axioms is a schema through which we are able to reason about the set of all natural numbers. Ordinality is implicit in the logical end results of these axioms in that each number has a unique successor in the set.

1.8.2 Addition

Although a rich literature indicates that non-human animals share an ability to order and compare numerosities with humans, less is known about other arithmetic operations. There is limited evidence for addition and subtraction in non-human animals. Addition is an arithmetic operation that involves combining numerical representations (addends) to form a new representation (the sum). The ability to combine addends nonsymbolically has been demonstrated in the spontaneous behavior of nonhuman primates and well as through training and transfer of a numerical rule. A limited, but strongly persuasive test of arithmetic in a nonhuman animal was conducted on a single, symbol-trained chimpanzee. In Boysen and Berntson (1989), experimenters hid several sets of oranges across multiple locations. The chimpanzee was able to
repeatedly select the Arabic numeral that corresponded to the sum of hidden sets of oranges for sets that summed to less than four items.

There is evidence that nonhuman animals spontaneously perform approximate addition on numerical values. Several studies have demonstrated that nonhuman primates reliably choose the larger of two food quantities by tracking one-by-one additions to multiple caches (e.g. e.g. Beran, 2004; Beran, Evans, Leighty, Harris, & Rice, 2008; Hanus & Call, 2007; Hauser, et al., 2000; vanMarle, et al., 2006). These studies provide evidence that one-by-one accumulation of items can be represented as their sum. However, stronger evidence for addition on approximate numerical representations requires that nonhuman animals be able to combine two sets.

Flombaum, Junge, and Hauser (2005) demonstrated spontaneous addition in rhesus macaques using a violation of expectancy paradigm. Untrained subjects saw an array of 2-4 lemons which was subsequently hidden by a screen, followed by an addition of 1-4 more lemons to the occluded amount, with possible outcomes of 4 and 8. Monkeys who saw an unexpected number of lemons (4 instead of 8, or 8 instead of 4) when the screen was removed looked longer than those who saw the expected number. However, when tested with the more difficult ratio 2+2= 4 or 6, the macaques who saw the numerically impossible outcome (6) failed to look longer than those who saw the numerically possible outcome.
Cantlon and Brannon (2007a) compared the nonverbal arithmetic abilities of monkeys and adult humans using a computer task in which they had to add the numerical values of two static sets of dots separated by a delay. After the presentation of these two sets, monkeys chose between two arrays: one numerically equal to the sum of the two sets and a distractor array. The addition problems consisted of addends ranging from 1 to 17. Cantlon and Brannon (2007a) reported that both humans’ and monkeys’ accuracy was ratio-dependent, suggesting that they performed addition by combining analog magnitude representations. Cantlon, Merritt and Brannon (under review) found a similar result using videos of moving dots. Altogether, these findings support the claim that approximate addition is an evolutionarily widespread cognitive process.

1.8.3 Subtraction

In contrast, the evidence for nonsymbolic subtraction by nonhuman animals is more complicated. Sulkowski and Hauser (2001) tested monkeys’ ability to subtract using a violation-of-expectation paradigm adapted from the infant literature. Subjects were shown an empty container into which a small number of eggplants were placed and/or removed. After presentation the contents of the container were revealed to the subjects. In a 2-1 subtraction condition, two eggplants were placed inside the container, and then one eggplant was removed. In a 1+1 addition condition, one eggplant was placed inside the container and another was added to it. As predicted by a violation of
expectancy, monkeys who saw the addition condition looked longer at the unexpected outcome of one eggplant. In the subtraction condition, however, although the majority of monkeys looked longer at the surprising outcome, the difference in looking time between surprising and unsurprising outcomes was not significant.

Beran (2004) reported a similar finding with spontaneous subtraction: chimpanzees watched as food items were added to or removed from two containers and were allowed to approach one container and consume the food items within. Chimpanzees chose the larger quantity when food items were added to containers but they were not successful at selecting the larger quantity when items were subtracted. The overall implication is that non-human animals’ capacities for subtraction are limited.

Cantlon, Merritt and Brannon (under review) used a training paradigm to provide evidence for subtraction abilities in monkeys. Monkeys were trained on a limited set of addition and subtraction problems. When tested with a nondifferentially reinforced novel problem set, they showed transfer of performance that allowed them to solve novel subtraction problems. Additional analyses ruled out simple heuristics such as selecting the choice option closest in numerosity to the largest, first, or second operand. Thus it appears that the subtraction operation are possible with training, but may not be understood spontaneously.
1.8.4 Multiplication

Perhaps the most complicated arithmetic ability that has been proposed to account for nonhuman animal behavior is multiplication. Multiplication primarily found to rely on rote memorization for adult humans (e.g. Dehaene & Cohen, 1997). This algorithm for memorizing symbolic multiplication facts is unique to humans. However, the act of combining magnitudes multiplicatively may have evolutionarily ancient roots.

Rate is number divided by time. Models of animal conditioning and foraging often include the assumption that animals can estimate or calculate rates of reinforcement. Leon and Gallistel (1998), for example, investigated whether the subjective magnitudes of rewards combined multiplicatively with the subjective rates of reward to determine a rat’s preference, as measured by the ratio of the time spent pressing one lever to the time spent pressing the other. Preference was determined both by rates of reinforcement and by the magnitudes of the rewards, and that the factor by which preference increased in response to a change in the rate was independent of magnitudes. These results indicate a combinatorial process that is a form of multiplication.

Measures of foraging time among various foraging sites also implicate an estimation of the relative rates of food occurrence (e.g. Baum & Rachlin, 1969; Herrnstein, 1961; Herrnstein & Loveland, 1975). Attempts to use a heuristic to model
this behavior without a calculation of rate have proved difficult (Lea & Dow, 1984).

However, a simple model can elegantly account for foraging behavior, a central
postulate of which is that an animal’s rate of switching to a foraging patch is determined
by the rate at which food has been observed to occur at that patch (Myerson & Miezín,
1980). Such research indicates that animals may have an ability to calculate rate, and
may do so using arithmetic information similar to multiplication.

There is preliminary evidence in human children which indicates that the
approximate number system may support multiplication. Barth, Baron, Spelke, and
Carey (2009) and McCrink and Spelke (2010) found that young children are able to
mentally transform numerosities multiplicatively. In both studies, children watch as a
visual array of objects becomes occluded. During occlusion, they are told that the array
is transformed by a “magic wand”, which changes the number of objects: doubling,
quadrupling, or increasing the numerical value by a fractional factor (2.5). Children are
then asked to compare their expectation of the outcome to a separate, visible array in an
ordinal discrimination. Performance on this task is ratio dependent, indicating that this
mental multiplication is performed on an analog magnitude representation of
numerosity.

Beran, Beran, Harris and Washburn (2005) provided limited evidence for
nonhuman primates’ ability to mentally double numerical sets. Two chimpanzees and a
rhesus macaque were trained on an association between 5 colors of containers and the
number of food items associated with each color, such that color became a cue for number. Subjects were able to compare pairs of same-colored containers with sets of visible food items. To successfully make this comparison, subjects had to mentally sum the contents of two identically colored containers, a process similar to doubling the numerical value associated with that color. However, in this study, color could serve as a symbol for number, and thus behavior could be adequately explained through a sort of symbolic addition process rather than multiplication.

To a similar end, Addessi, et al. (2008) trained capuchin monkeys to trade tokens of different values for food. Like color, tokens act as symbols by arbitrarily representing numerical value without having any iconic relation to their referent. Subjects were able to choose between combinations of tokens A and B, worth one and three rewards, respectively. Selecting the larger reward amount required that subjects estimate the number of tokens, representing what value each token stood for compute a total arithmetically. Capuchins were able to flexibly combine quantities to maximize their reward. However as with Beran, et al. (2005), it is unclear whether subjects represented a single set and doubled that value mentally, or whether the use of tokens led to symbolic addition.

1.9 Research Approach

The studies presented here take a comparative, behavioral approach to characterizing species differences and similarities in the approximate number system,
and the contexts that affect that system. While task parameters differ, the following studies generally focus on the ability to form approximate numerical representations of visual sets, and compare those representations, while adapting the task to address context, precision, and species comparisons. Chapter 2 will address quantitative differences in the acuity of the ANS between multiple primate species, using a computerized task with identical parameters. Chapter 2 further explores the effect of training on the precision of numerical representations. Chapter 3 is concerned with qualitative similarities between the numerical comparison process of humans and that of monkeys. Specifically, it addresses similarities in the effect of contextual change on numerical comparisons. Chapter 4 tests the hypothesis that an object tracking system, rather than the ANS, is invoked in the context of spontaneous numerical judgments for nonhuman primates as it is for human infants. Finally, Chapter 5 explores the flexibility of the ANS in supporting an arithmetic operation, approximate multiplication, using a conceptual model proposed to be implicit to humans’ understanding of multiplication.

A large literature indicates that the ability to represent number is widespread throughout the animal kingdom. By and large, the focus in the field has been on providing existence proofs of numerical competence in wide-ranging taxa or using individual primate species as models for comparisons with humans using a wide variety of tasks (for reviews, see Beran, 2008b; Brannon, 2006). Here we take a different approach by systematically comparing multiple species with a single task and set of stimulus parameters in an attempt to provide a more quantitative assessment of species differences and similarities in numerical cognition.

Including prosimian primates in parametric comparisons of primate cognition is important in attempting to identify cognitive profiles of the primate ancestral state. The prosimian suborder of primates (lemurs, lorises and galagos) diverged from other primates approximately 50 million years before monkeys diverged from humans (75 vs. 25 million years ago; Horvath, et al., 2008). Prosimians are considered to be among the most “evolutionarily conserved” primates, in so far as they have been hypothesized to be morphologically and behaviorally similar to the last common primate ancestor (Tattersall, 1982; Yoder, 2007). If prosimians share cognitive traits that are common
among other primates, it is likely that these traits were present in the last common ancestor.

Much less is known about cognition in lemurs than in monkeys or apes. The limited data available, however, seem to indicate that lemurs differ from apes and monkeys both quantitatively and qualitatively on measures of social intelligence. For example, lemurs do not exhibit within group coalitions, deception or elaborate visual communication (for a review, see Fichtel & Kappeler, 2010). Monkeys outperform lemurs in visual co-orientation with humans (Anderson & Mitchell, 1999; but see Shepherd & Platt, 2008; Ruiz, Gomez, Roeder & Byrne, 2009) and vervet monkeys have been found to be superior to ring-tailed lemurs (and 1 brown lemur) on a reversal discrimination task (Rumbaugh & Arnold, 1971).

On the other hand, a number of recent studies have found qualitative similarities between lemurs and monkeys in non-social cognitive tasks. For example, lemurs perform more like macaque monkeys than pigeons in a sequence learning task (Merritt, McLean, Jaffe & Brannon, 2007). Lemurs are capable of transitive reasoning (MacLean, Merritt & Brannon, 2008), Piagetian object displacement (Deppe, Wright & Szelistowski, 2009), and learning a reverse contingency in a food reward task (Genty, Palmier & Roeder, 2004). Lemurs also perform similarly to monkeys on assays of spatial memory (Erhart & Overdorff, 2008; Lührs, et al., 2009), inferring causality (Santos, Barnes &
While a large number of studies have demonstrated that monkeys and apes can represent number abstractly (e.g., Brannon & Terrace, 1998; Beran, 2004), only a few studies have investigated numerical cognition in prosimian primates (Lewis, Jaffe & Brannon, 2005; Merritt, MacLean, Crawford & Brannon, 2011; Santos, Barnes & Mahajan, 2005). Santos, Barnes, and Mahajan (2005) investigated the spontaneous representation of number in lemurs (brown lemurs, mongoose lemurs, ring-tailed lemurs, and ruffed lemurs) with a modified version of the Wynn (1992) violation of expectancy paradigm. Lemurs watched as two lemons were hidden behind a screen; the screen was then lowered to reveal either a possible outcome (two lemons), or an impossible outcome (one lemon or three lemons). On average, lemurs looked longer at the impossible outcome than at the possible outcome, suggesting an ability to represent the number of lemons behind the screen. Lewis, Jaffe, and Brannon (2005) modified another task designed for human infants (Feigenson & Carey, 2003) to explore numerical abilities in lemurs. In this study, mongoose lemurs watched as grapes were hidden one at a time in a bucket that was filled with shredded paper. A subset of the food items were hidden in the container’s false bottom. Lemurs were allowed to retrieve and consume the available food items and additional search time was then measured. If lemurs expected additional grapes to be in the bucket they should have continued to manually search
after retrieving the available food items. Results indicated that search time was dependent on the number of grapes that should have been in the bucket and reflected ratio dependent discrimination (e.g., lemurs searched longer after retrieving 2 grapes if 4 grapes had been hidden, a 1:2 ratio of retrieved grapes to presented grapes, but they did not search longer after retrieving 2 grapes if 3 grapes had been hidden).

In contrast to the Santos, Barnes and Mahajan, (2005) and Lewis, et al., (2005) studies which used food items as the discriminative stimuli and examined spontaneous numerical discrimination, Merritt, MacLean, Crawford, and Brannon (2011) used a touch-screen task and trained two ring-tailed lemurs to order visual arrays based on numerosity. Lemurs were trained to select the smaller of two arrays of dots (values 1 to 4), using trial-unique stimuli and surface area controls. After successfully learning to order pairs of the values 1-4, the lemurs were able to extrapolate this ordinal rule to novel numerical values (arrays of 5 to 9 dots) replicating the basic pattern of results found by Brannon and Terrace in macaque monkeys (Brannon & Terrace, 1998, 2000). While lemur performance in this ordinal numerical comparison task was qualitatively similar to that of monkeys, accuracy was generally lower for the ring-tailed lemurs compared to the macaques. It is unclear, however, whether these differences reflect true species differences in capacity or instead reflect task parameters and training history (e.g. Savage & Snowdon, 1989).
Lemur performance in the Merritt, et al., (2011) study showed the hallmark signature of the approximate number system (ANS), wherein numerical discrimination was modulated by ratio in accordance with Weber's law. According to Weber's law, the increase in intensity required to detect a change to a stimulus is a function of the magnitude of the initial intensity. Specifically, the numerosity of a set of n dots is represented internally by a Gaussian random variable. The internal weber fraction, w, specifies the precision of the internal quantity representation. However, it is also important to note that weber fractions are not hard and fast invariants. The absolute value will vary based on the model used to estimate w and may also vary based on a multitude of different factors such as the duration of stimulus presentation or the range of numerical values used. These factors mean that it is essential to use the same task parameters when attempting to make species comparisons.

We tested three species of lemurs (ring-tailed, mongoose and blue-eyed black) and rhesus macaques in an identical ordinal touch-screen task to more quantitatively assess species differences in weber fraction for numerical discrimination. Prosimian primates are a suborder of primates with tremendous diversity in social structure and ecology. Despite their close phylogenetic relatedness, prosimians primates differ widely in their home range, social group size, and diet. By directly contrasting the performance of lemur species that differ in these traits, we are able to test the hypothesis that differences in ecology or social structure cause large differences in non-social
cognition. To this end, we tested three lemur species: blue-eyed black lemurs, ring-tailed lemurs, and mongoose lemurs. These three species differ with respect to diet, social structure and home range size. If selective pressure to discriminate between quantities of discrete food items results in more precise number representations, mongoose lemurs and blue-eyed black lemurs would be expected to outperform ring-tailed lemurs (opportunistic omnivores) by virtue of a dietary reliance on fruit. Alternatively, if selective pressures to discriminate between the number of conspecifics in one’s group and/or a rival group results in greater numerical acuity, ring-tailed lemurs (and to a lesser extent, blue-eyed black lemurs) would be expected to outperform mongoose lemurs. Finally, if spatial abilities associated with a large home range result in a more evolved magnitude representation system, the broad ranging ring-tailed lemurs should out-perform other species.

A second question our research addressed relates to the malleability of the ANS. Research with human subjects demonstrates that the ANS matures gradually over development and continues to sharpen into late adolescence (Halberda, Mazzocco, & Feigenson, 2008). There is less research addressing the cause of this sharpening. Here we ask whether extensive practice improves the acuity of the ANS.
2.1 Experiment 1A

2.1.1 Methods

2.1.1.1 Subjects

Subjects were 4 male individuals from each of four species: mongoose lemurs (*Eulemur mongoz*, mean age 18.68 years, standard deviation 5.92), ring-tailed lemurs (*Lemur catta*, mean age 8.41 years, standard deviation 6.05), blue-eyed black lemurs (*Eulemur macaco flavifrons*, mean age 12.30 years, standard deviation 3.95) and rhesus macaques (*Macaca mulatta*, mean age 9.21 years, standard deviation 4.10). All animals were experimentally naïve. Animals were neither food- nor water-restricted.

2.1.1.2 Apparatus

Prosimian subjects were housed in indoor enclosures at the Duke University Lemur Center (DLC). Animals were housed either singly or in pairs and those housed in pairs were either separated from their cage-mate during testing or their cage-mate was trained not to interfere. A custom-built, stainless steel testing station (86cm x 43cm x 35cm) containing a 15-inch touch-sensitive computer monitor, a desktop computer and a reward delivery system was used to test prosimian primates. The apparatus was wheeled into each subject’s home enclosure for all training and testing sessions. A clear Plexiglas panel with circular openings (diameter 5 cm) over each stimulus location covered the screen to prevent unnecessary contact with areas of the touchscreen that did
not contain stimuli. Lemurs were trained to sit on a small plastic crate in front of the cart to engage in the task.

Macaques were housed either singly or in pairs in a vivarium. Those housed in pairs were separated for testing. All animals had been trained to use the touch screen for a previous task, but had no numerical discrimination training. A 15-inch touch-sensitive computer monitor and a reward delivery system were attached to the front of each macaque’s cage for each test session. Macaques were kept on water-restricted diets for participation in other tasks, as approved by an Institutional Animal Care and Use Committee.

2.1.1.3 Task

The experimental program was identical for the four species. A custom-built REALbasic program displayed the stimuli and recorded responses. Subjects initiated each trial by touching a red square presented in the bottom center of the screen. Following the start-response, two yellow squares, each containing an array of red dots, were presented in two central screen locations. The location of the correct stimulus (left vs. right) was determined randomly. Subjects were rewarded for selecting the array containing the larger number of dots. Correct responses elicited positive visual and auditory feedback and a food reward. Incorrect responses elicited a warning tone and a black screen for 6 s. All trials were followed by a 3 second inter-trial interval (ITI). For lemurs, correct responses were rewarded with a 190 mg fruit punch-flavored TestDiet®
pellet (Division of Land O’Lakes, Purina Feed, LLC, Richmond, IN, U.S.A.), and for macaques the same fruit flavored pellet or a miniature M&M, or Reese's Pieces candy.

2.1.1.4 Stimuli

Stimuli were 200x200 pixel yellow squares within which 1-9 elements were randomly arranged (Figure 1). The computer program randomly selected the size and location of the dots in each array and generated trial unique images with the restriction that dots did not overlap with other dots or with the boundary of the square in which they were contained. To ensure that the task could not be solved by attending to the cumulative surface area of the arrays or the size of the individual elements, the physical sizes of the elements were varied such that on 50% of trials, element size was equated between the two choice stimuli and on the other 50% of trials, cumulative surface area was equated. The cumulative surface area as well as the individual element size of the choice stimuli varied between 174 and 7,854 pixels.

Figure 1: Example stimuli in ordinal task.
2.1.1.5 Procedure

Subjects were tested Monday through Friday for one session per day. Two individuals of each species were trained with the pairs {1 vs. 9, 1 vs. 3, 3 vs. 9}, and the other two individuals were trained with the pairs {2 vs. 8, 2 vs. 4, 4 vs. 8}. Each training session contained 20 trial-unique sets per numerosity pair each presented 3 times for a total of 60 trials. Each subject was trained until accuracy reached 75% or higher for two sessions, or until 30 training sessions had been completed. After the criterion was met, each animal moved on to the test phase of the experiment. All the animals were tested with all 36 numerosity pairs made possible by combining the numerosities 1-9. Lemurs were tested with 15 72-trial sessions and monkeys were tested with 30 36-trial sessions.

2.1.1.6 Analysis

Numerical sensitivity was quantified in three ways for each individual. First, for each individual, we calculated the just noticeable difference (JND), defined as the ratio at which a linear trendline fit to a subject’s data crossed the 75% accuracy level. Second, we estimated each subject’s weber fraction using the model presented by Pica, Lemer, Izard and Dehaene (2004), hereafter referred to as $w_{Pica}$. The Pica model assumes that the numerosity of a set of n dots is represented internally by a Gaussian random variable. This model includes one free parameter: the internal weber fraction w. Finally we estimated each subject’s weber fraction using a modification of the model presented in Piazza, Izard, Pinel, Le Bihan and Dehaene (2004), hereafter referred to as
The modification reflects task differences: the Piazza model assumes subjects have an exact representation for some standard value; w is calculated using the overlap of distribution functions of internal representations for each test value with the exact standard representation. We assume that nonhuman primates do not represent any values exactly, thus both values are represented approximately. The model was therefore modified to calculate w using the overlap of the distribution function of internal representations for the two numbers being compared. The Piazza model assumes a logarithmic internal number line with fixed Gaussian variability (although, note that a linear number line with scalar variability predicts virtually identical results). This model includes a single free parameter: the internal weber fraction w.

We performed a hierarchical Bayesian analysis (Wagenmakers, Lodewyckx, Kuriyal & Grasman, 2010) using Markov Chain Monte Carlo (MCMC) simulation via the RJAGS package in R\textsuperscript{1}. For the data reported in Table 1, we assumed the w parameter for each individual to be drawn from a gamma distribution, with species-specific mean and variance also drawn from gamma distributions with scale and rate parameters (4,3) and (3,4), respectively. Results were remarkably robust across testing with alternative priors, as documented in Online Resource 1. We calculated posterior distributions for individual w’s and species-specific means and standard deviations by simulating pairs.

\textsuperscript{1} All Bayesian analyses are reported here as published in Jones, S. M., Pearson, J., DeWind, N., Paulsen, D., Tenekedjieva, A. & Brannon, E.M. (under review). J. Pearson and A. Tenekedjieva conducted all Bayesian analyses, and analyses are included here with permission.
of 20,000 sample Markov chains (thinning fraction=3) for each model, preceded by 5000 iterations of burn-in. We verified chain convergence via examination of autocorrelations and Geweke's diagnostic, implemented via the CODA package in R.

To compare models, we calculated Bayes factors for four models: 1) Pica model with species-specific distributions of \( w \), 2) Pica model with a single, cross-species distribution of \( w \), 3) Piazza model with species-specific distributions for \( w \), and 4) Piazza model with a single, cross-species distribution of \( w \). While a variety of approaches to calculating Bayes factors exist (e.g. Kass & Raftery, 1995; Wagenmakers, et al., 2010), we used the Laplace-Metropolis method (see Online Resource 1), which is a good approximation for large datasets such as ours. In interpreting these results, we followed the recommendations in Kass and Raftery (1995), in which \( 2\log(BF) \) gives evidence in favor of the experimental hypothesis that is considered "positive" for values in \([2,5]\), "strong" in \((5,10]\), and "very strong" for \( 2\log(BF)>10 \).

### 2.1.2 Results

#### 2.1.2.1 Training

Twelve of the fifteen animals met the criterion on the training values in fewer than 30 sessions. There was a large amount of inter-subject variability in the number of trials needed to reach criterion (range = 411 and 2204 trials). Two blue eyed black lemurs never reached criterion and were moved on to testing after 30 training sessions. Due to experimenter error, a third blue eyed black lemur was moved on to testing after 8
training sessions (480 trials), despite having only reached 75% once rather than twice for the pair 1 v. 3.

A 2 (training values: {2,4,8} vs. {1,3,9}) X 4 (species) ANOVA on number of trials to reach criterion revealed no significant main effect of species, $F(3,15) = 2.16$, NS, or training values, $F(1,15) = 0.66$, NS. There was, however, a significant interaction between species and training values, $F(3,15) = 4.77$, $p < 0.05$. This interaction reflects the fact that for two species (mongoose lemurs and rhesus macaques) animals trained on the values 1, 3 and 9 required more trials to reach criterion than the animals trained with the values 2, 4, and 8. In contrast the other two species (blue-eyed black lemurs and ring-tailed lemurs) showed the opposite pattern. A larger sample size would be necessary to determine if this interaction reflects meaningful species differences or instead individual differences.

2.1.2.2 Testing

Each of the 16 subjects selected the larger value more often than predicted by chance in test (all $p$ values $< 0.001$). Average accuracy across test sessions ranged from 59% to 91% correct ($M = 72\%$, $SD = 0.08$). To compare average accuracy across the four species, we conducted a one-way ANOVA. Results revealed no significant main effect of species, $F(3,12) = 0.105$, NS.
2.1.2.3 Ratio-dependent performance

Like the macaques and ring-tailed lemurs in Merritt, et al. (2011), all four species showed a clear effect of ratio on accuracy, consistent with Weber’s law (Figure 2). Accuracy decreased as the ratio between the two numerosities being compared approached one, \( r(26) = -0.98, p < 0.001 \), for testing with pairs 1-9. In contrast, we found no relationship between ratio and median response time for correct trials, \( r(26) = -0.01, \) NS.

Figure 2: Accuracy as a function of numerical ratio from Experiment 1. Unlike the results of Merritt, MacLean, Crawford & Brannon (2011), we found no prominent difference in accuracy between lemurs and monkeys.

---

2 The lack of relationship between ratio and RT in this dataset may be due to the fact that the lemurs responded to the touchscreen using their nose, which led to much slower and more variable RTs than those of the macaques or human participants. In addition the lemurs were in large enclosures with conspecifics competing for their attention.
2.1.2.4 Quantifying numerical sensitivity

All three measures of numerical sensitivity were strongly correlated: $w_{\text{Pica}}$ and $w_{\text{Piazza}}$ were significantly correlated, $r(14) = 1.00, \ p < 0.001$, $w_{\text{Pica}}$ and JND and $w_{\text{Piazza}}$ and JND were both significantly negatively correlated, $r(14) = -0.97, \ p < 0.001, \ r(14) = -0.97, \ p < 0.001$, respectively.

JND in this case provides a directly comparable empirical measure, which allows for a comparison of the data reported in this study with previously reported numerical discrimination abilities in primates. Two animals, the worst performing blue-eyed black lemur and the worst performing mongoose lemur, were excluded from this calculation as their JND was a negative number, indicating poor accuracy overall. For all other animals, a linear regression provided R2 values between 0.63 and 0.78. JNDs for ring-tailed lemurs and blue-eyed black lemurs ranged from approximately a 1:3 ratio to a 1:2 ratio. JNDs for mongoose lemurs and macaques ranged from approximately a 1:4 ratio to a 1:2 ratio. This indicates that while some individuals within these species may be able to discriminate a 1:2 ratio approximately 75% the time, as has been reported in other studies (e. g. Lewis, et al., 2005), there are some individuals in each species which require easier ratios to achieve 75% accuracy on numerical discriminations.

To further quantify uncertainty in our fits of $w_{\text{Pica}}$ and $w_{\text{Piazza}}$ and perform model comparisons, we performed a hierarchical Bayesian analysis (Wagenmakers, et al., 2010). In our case, Bayesian analysis has the advantage of incorporating our prior intuitions.
about the size and distribution of w's for different individuals in a principled way, and potentially offering support for (not mere rejection of) the null hypothesis. The hierarchical Bayesian analysis provides a quantification of the strength of the evidence for the hypothesis of no species difference. Table 1 shows log differences in Bayes Factors, a measure of strength of evidence for each of the models. Species-specific models are less preferred than those positing a single distribution of w's across all species for each choice model, with the two Pica models outperforming the two Piazza models. Thus, the Bayesian analysis strongly supports the hypothesis that there are no species differences and that instead the 16 data points were pulled from a single distribution of w values (see Figure 3 for individual weber fractions by species). The model also favors the Pica model as a description of numerical encoding and choice.

When calculating small weber fractions within a relatively small range, the Pica and Piazza models perform equally well (as in Experiment 2). However, when there is a wider range of larger weber fractions, such as those seen in Experiment 1, the Pica model has a higher log likelihood than the Piazza model (Table 1).

Log Bayes Factors for comparing any two models may be calculated as twice the difference between the relevant columns, with positive values favoring the first model and negative values the second. Log Bayes factors of >10 are considered strong evidence, while those <2 are considered negligible. Thus, in experiment 1A, models in which w's are drawn from species-specific distributions are less preferred than those positing a
single distribution of w’s across all species, with the two Pica models outperforming the
two Piazza models. In contrast, in experiment 1B, results dramatically favor models in
which w’s are drawn from species-specific distributions, indicating that the underlying
distribution of weber fractions in humans differed significantly from the underlying
distribution weber fractions for nonhuman. Finally, the results of experiment 2 indicate
an effect of extensive numerical experience: models in which w values for the two
macaque samples (experience limited from Experiment 1A vs. extensive experience from
Experiment 2) are drawn from distinct distributions out performed models in which all
w values were drawn from the same distribution.

Table 1: Model fits for Chapter 2, Experiments 1A, 1B, and 2. Differences
from maximum log likelihood (best fit) for each of four models.

<table>
<thead>
<tr>
<th>Model Fits - strength of evidence for each of four models</th>
<th>Experiment</th>
<th>Linear model (Pica), no difference between distributions</th>
<th>Linear model (Pica), different distributions</th>
<th>Log model (Piazza), no difference between distributions</th>
<th>Log model (Piazza), different distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1A</td>
<td>0</td>
<td>-5.39</td>
<td>-21.35</td>
<td>-25.46</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>-6.96</td>
<td>0</td>
<td>-31.63</td>
<td>-20.09</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-571.7</td>
<td>$&lt;$ 10$^5$</td>
<td>-615.6</td>
<td>0</td>
</tr>
</tbody>
</table>

A major limitation of our method for species comparisons is that we were only
able to test four individuals per species, which, although a typical sample size for
comparative primate research, is unlikely to provide enough power to detect small
species differences. We directly tested this by performing two types of power analysis. First we asked how many individuals would be required for the observed differences in mean w's across species to be significant at the \( p < 0.05 \) level. These numbers ranged from as few as 36 (between *Lemur catta* vs. *Eulemur mongoz*) to as large as 3200 (*Lemur catta* vs. *Macaca mulatta*). This analysis assumes that the larger sample would maintain the same mean and variance observed in the small samples that we obtained for each species. Our second approach was to ask how large the difference in w between species would have to be, given our sample sizes, to be detected with a 10% chance of Type II error (false negative). This analysis suggested that the difference in w would need to be larger than 0.53 in the best case (*Lemur catta* vs. *Macaca mulatta*) and larger than 1.43 in the worst case (*Lemur catta* vs. *Eulemur mongoz*) to be detectable with our sample size.
Figure 3: Weber fractions ($w_{\text{Pica}}$) for the three lemur species and rhesus macaques from Experiment 1A, adult humans from Experiment 1B and the three extensively trained rhesus macaques from Experiment 2. Black circles represent medians of posterior distributions for individual weber fractions. Boxplots show quantiles of posterior distributions for species specific weber fractions in the Pica model. Distance between whiskers represents the 95% credible interval for the species-specific $w$.

2.1.2.5 Surface area

To determine whether subjects were using number or cumulative surface area to order the stimuli, we conducted separate analyses on trials when 1) cumulative surface area was equated across both stimuli, and 2) dot size was equated across both stimuli. Overall, we found that accuracy was above chance regardless of whether the cumulative
surface area of the stimuli was equated (M = 70.04%, t(15) = 14.01, p < 0.0001) or the size of the elements in each stimulus was equated (M = 70.45%, t(15) = 13.76, p < 0.0001). We calculated a surface area difference score for each animal (\(w_{Pica}\) for trials in which cumulative surface area was equated across both stimuli subtracted from \(w_{Pica}\) for trials in which dot size was equated). On average, difference scores did not differ significantly from zero, \(t(15) = 0.57, NS\), indicating no consistent effect of cumulative surface area congruence on weber fraction across subjects.

To confirm that this pattern did not differ by species, we conducted a one-way ANOVA. Results revealed no significant main effect of species, \(F(3,12) = 0.425, NS\). One macaque and one mongoose lemur, however, had large positive difference scores (0.37 and 0.29, respectively) indicating that their performance was superior when cumulative surface area was congruent with numerosity. In addition, two macaques and one blue-eyed black lemur had large negative difference scores (macaques: -0.28, -0.26; blue-eyed black lemur: -0.49), indicating that their performance was superior when surface area for the two numerosities compared was equal.

To assess the changing effect of surface area over time, we calculated a difference score for the first and last block of 360 trials of testing for each individual. A Wilcoxon signed-rank test revealed that there was no significant difference between difference scores for the first and last testing block; \(z = 0.71, NS\), indicating that the effect of surface area congruence did not change over the course of testing.
2.1.2.6 The effect of training on numerical acuity

To assess the effect of experience on numerical sensitivity we compared \( \nu_{Pica} \) for the first and last block of the three 360 trial-test blocks. A Wilcoxon signed-rank test revealed that there was a significant difference between \( \nu_{Pica} \) for the first and last testing block; \( z = -2.07, p < 0.05 \). Thirteen of the 16 individuals showed a decrease in weber fraction from the first block of testing to the last block of testing indicating an improvement in numerical discrimination over repeated testing.

2.2 Experiment 1B

The goal of Experiment 1B was to provide a clean comparison between humans and nonhuman primates. While a number of papers have reported weber fractions for adult humans in similar tasks (e.g. Cantlon & Brannon, 2006b; DeWind & Brannon, 2012; Gilmore, Attridge & Inglis, 2011; Halberda, Mazzocco, & Feigenson, 2008; Piazza, et al., 2004; Pica, et al., 2004) we conducted Experiment 1B to assess weber fraction with a procedure as close as possible to Experiment 1A and with the identical stimuli to allow a quantitative comparison between humans and nonhuman primates.

2.2.1 Methods

2.2.1.1 Subjects

Ten adult human participants were tested with the same descending ordinal task used in Experiment 1A (mean age 24.51 years, SD 10.58, 4 males). Data for each
participant was collected in a single 1-hour session. All participants were compensated $10 for completion of the task.

2.2.1.2 Task, stimuli and procedure

Pilot data revealed that when humans were tested with unlimited time to make a response their performance was at ceiling and thus their weber fraction could not be calculated. Consequently, stimulus pairs were presented for 100 ms. The task was otherwise identical to that described in Experiment 1A with the following minor exceptions: a) no auditory feedback or food rewards were provided, b) Stimuli were presented for 100 ms and then replaced with gray 200 x 200 pixel placeholders, c) responses were made via a mouse click on the gray square that replaced that array. These modifications serve to make the task for humans more difficult than the task for nonhumans. Thus, this comparison is likely to minimize differences between human and nonhuman primates. Finally, d) participants were tested on a single session of 406 to 648 trials with pairs of the numerosities 1 through 9 and they were verbally instructed to select the larger numerical value.

2.2.2 Results

All participants selected the larger value more often than predicted by chance (all \( p \) values < 0.001). Overall accuracy ranged from 87% to 95% correct (M = 92%, SD = 0.03). There was a clear effect of ratio on accuracy, consistent with Weber’s law. Accuracy decreased as the ratio between the two numerosities being compared approached one,
$R^2 = -0.84$, $p < 0.001$. In contrast to the results for nonhuman primates in Experiment 1A, we found a significant positive correlation between ratio and response time for correct trials, $R^2 = 0.92$, $p < 0.001$.

2.2.2.1 Quantifying numerical sensitivity

Participant’s $w_{Pica}$ ranged from 0.16 to 0.33 ($M = 0.21$, $SD = 0.05$) (Figure 3). Using a hierarchical Bayesian analysis to quantify the strength of the evidence for the hypothesis of no species difference between the nonhuman primates in Experiment 1A and the humans in Experiment 1B, we found that the species-specific models dramatically outperformed models that posit a single distribution of $w$'s across both human and nonhuman primates. In addition, the Pica model outperformed the Piazza model. Thus, the Bayesian analysis strongly supports the hypothesis that the human $w$ values were drawn from a different distribution of $w$ values than the nonhuman primates tested in Experiment 1A.

2.2.2.2 Surface area

Accuracy was above chance regardless of whether the cumulative surface area of the stimuli was equated ($M = 91.40\%$, $t(9) = 37.81$, $p < 0.0001$) or the size of the elements in each stimulus was equated ($M = 93.33\%$, $t(9) = 63.93$, $p < 0.0001$). A surface area difference score was calculated for each participant as in Experiment 1A ($w_{Pica}$ for trials in which cumulative surface area was equated across both stimuli subtracted from $w_{Pica}$ for trials in which dot size was equated). On average, difference scores differed
significantly from zero, \( t(9) = 2.37, p < 0.05 \), indicating that participants had lower weber fractions when cumulative surface area was congruent with numerosity then when cumulative area was equated.

2.2.3 Discussion

In summary, in Experiment 1A we found no species difference between macaques and lemurs in numerical acuity as measured by \( w \) derived from a numerosity ordering task. This differs from a previous finding by Merritt, et al. (2011), in which macaques were found to outperform lemurs on the same type of task. The discrepancy in conclusions highlights the importance of using identical task parameters and training procedures in comparative research. In addition, our results indicate that the Pica model outperforms the Piazza models when there is a wider range of larger weber fractions.

In contrast, comparing the results of Experiments 1A and 1B demonstrates a clear species difference between humans and nonhuman primates, indicating significantly better numerical acuity in adult humans compared to macaques or lemurs. This was true despite the modifications to the procedure that we used to avoid a ceiling effect in humans, indicating that the difference between the numerical acuity of adult humans and the numerical acuity of macaques and lemurs is likely greater than reported here.

A secondary finding from Experiment 1A was that subjects’ numerical acuity markedly improved between the first and last block of testing. This raises the question
of how malleable the approximate number system is and whether numerical acuity would continue to improve over extensive practice. Experiment 2 addresses this question.

### 2.3 Experiment 2

Experiment 2 tested three rhesus macaques that had extensive prior training on numerical ordering on the same task and stimuli used in Experiment 1. We predicted that macaques with extensive numerical training would show lower Weber fractions compared with the macaques in Experiment 1 that had limited numerical training.

#### 2.3.1 Methods

**2.3.1.1 Subjects and housing**

Subjects were three female rhesus macaques (mean age 12.39 years, standard deviation 1.88) with extensive prior numerical discrimination training in a variety of tasks (e.g. Cantlon & Brannon, 2005, 2006a,b, 2007a,b; Jones, Merritt, Cantlon & Brannon, 2010; Jordan & Brannon, 2006a,b; Jordan, MacLean & Brannon, 2008). Feinstein had been trained on numerical ordering, matching and bisection tasks, and nonsymbolic addition and subtraction tasks, totaling approximately 375,200 trials. Mikulski and Schroeder had both been trained on numerical ordering, matching, and bisection tasks, totaling approximately 186,000 trials, and 208,000 trials respectively.

Animals were socially housed but were tested in isolation. Monkeys were rewarded with juice rather than food and were kept on water-restricted diets to increase
their motivation for juice reinforcement, as approved by an Institutional Animal Care and Use Committee.

2.3.1.2 Task, stimuli and procedure

The task and stimuli were identical to those used in Experiment 1A. Correct responses were rewarded with 1.3-ml squirt of Kool-Aid or water.

Subjects were tested Monday through Friday for 1 session per day in sound-attenuated booths while seated in Plexiglas chairs fitted with a liquid-delivery system. Unlike the subjects in Experiment 1, subjects were tested outside of their home enclosures. Stimuli were presented on a 17-in. touch-screen computer monitor fixed to the interior of the soundproof booth. All animals were tested with all 36 numerosity pairs made possible by combining the numerosities 1-9. Subjects were tested with the same total number of test trials as in Experiment 1, however they were tested in four 270-trial sessions rather than over 15 or 30 sessions.

2.3.2 Results and Discussion

Both accuracy and response time were modulated by ratio such that accuracy decreased ($R^2 = 0.76, p < 0.001$) and response time increased ($R^2 = 0.49, p < 0.0001$) as the ratio between the comparison numerosities approached 1. A Bayesian analysis using the same hierarchical models as Experiment 1A showed very strong evidence for the hypothesis that $w$ values for the two macaque samples (experience limited from Experiment 1A vs. extensive experience from Experiment 2) are drawn from distinct
distributions (Table 1). The results of this analysis must still be interpreted in light of the fact that only two of the three macaques showed higher numerical sensitivity compared to the macaques in Experiment 1A (Feinstein and Mikulski: $w_{\text{Pica}} = 0.25$ and $0.38$, respectively; Schroeder $w_{\text{Pica}} = 0.76$). Schroeder showed numerical sensitivity in the same range as the lesser-trained macaques and the lemurs (Figure 3). It is important to note that Schroeder had considerably less training on ordinal tasks than the other two monkeys with only one year of experience ordering numerosities compared to 6.67 and 4.75 years of prior ordinal training. It is unclear whether this difference in numerical experience accounts for Schroeder’s reduced performance relative to the other two experienced monkeys or whether her performance reflects more stable individual differences given that Schroeder has consistently performed poorly in comparison with other macaques in a variety of tasks (Cantlon & Brannon, 2007b; Jones, et al., 2010; Merritt, Rugani & Brannon, 2009).

### 2.4 Experiment 3

Experiment 2 suggests that extensive numerical training improves weber fraction. However given that only two of the three rhesus macaques with extensive training outperformed the four rhesus macaques with limited numerical discrimination training, we next sought to examine the effect of extensive training using a within-subjects design. Thus, in Experiment 3, we retested all three monkeys from Experiment 2 on a task identical to one they had been trained on several years earlier (before
Experiment 2). In the intervening years, they had extensive and varied numerical discrimination experience described below.

2.4.1 Methods

2.4.1.1 Subjects

The same three macaques used in Experiment 2 were tested in Experiment 3. We chose not to compare early training performance to the data collected in Experiment 2, because stimulus parameters were not identical. Consequently we retrieved data from early in their training (April and May of 2005 for Feinstein and Mikulski, and June of 2008 for Schroeder) and considered that data to be Time 1. We then retested the monkeys with the same stimulus parameters and numerical values used at Time 1 in September of 2011 and we label this new data set Time 2.

The Time 1 block of data was chosen as a basis for comparison because it represents early ordering experience for each monkey that is as closely matched as possible to the amount of training received by the limited-experience macaques and lemurs tested in Experiment 1. Prior to Time 1, all three monkeys were trained to order a subset of the numerosities 1-9. Feinstein had been trained with the numerosities 1, 3, 5, and 9. Mikulski was trained with the numerosities 1, 4, 5, and 9. Schroeder was trained with the numerosities 2, 4, and 8. After their initial training the numerical values were gradually expanded to include all possible pairings of the numerosities 1-9. Between Time 1 and Time 2, all three macaques received extensive training on numerical tasks.
Feinstein was trained on numerical ordering, matching and bisection tasks, and nonsymbolic addition and subtraction tasks, totaling approximately 367,700 trials. Mikulski and Schroeder were trained on numerical ordering, matching, and bisection tasks, totaling approximately 162,000 trials, and 61,000 trials respectively.

2.4.1.2 Task, stimuli and procedure

The data from Time 1 represent the first few sessions of training using all numerosities 1-9 (Feinstein: 1235 trials, presented over 4 sessions; Mikulski: 1254 trials, presented over 2 sessions; Schroeder 1200 trials, presented over 12 sessions). At Time 2 all three monkeys were tested for 1200 trials across 4-6 sessions.

The task was identical to that described in Experiment 1 and 2 with the following exceptions. The start-response icon was black rather than red, and located in the lower right hand corner of the screen rather than centrally. Nine rather than two spatial locations were used to present the two numerosity stimuli on each trial. In contrast to Experiments 1 and 2, subjects were rewarded for selecting the array containing the smaller number of dots followed by the array containing the larger number (ascending rather than descending numerical order). As in Experiment 2, correct responses elicited positive visual and auditory feedback and a juice or water reward. Incorrect responses elicited a warning tone as in Experiments 1 & 2 accompanied by a 2 second rather than a 6 second timeout. All trials were followed by a 1.5 second inter-trial interval (ITI) rather than the 3 second ITI used in Experiments 1 and 2. Stimuli were yellow squares
presented on a red rather than blue background. Stimulus controls also differed slightly from those in Experiments 1 and 2. In Experiment 1 & 2, element size was equal on 50% of trials and cumulative surface area was equal on the other 50% of trials. In contrast, the stimuli in Experiment 3 contained 50% of trials in which cumulative surface area was congruent with numerosity (the larger numerosity had the larger cumulative surface area) and 50% of trials in which cumulative surface area was incongruent with numerosity.

In contrast to the other monkeys, Schroeder was tested only using the numerosities 2, 3, 4, 6, 8, and 9. All trials were followed by a 1 second inter-trial interval (ITI), which differed from the 1.5 second ITI used for Feinstein and Mikulski. The stimuli in Experiment 3 contained 40% trials (vs. 50% for Feinstein and Mikulski) in which cumulative surface area was congruent with numerosity and 60% of trials in which cumulative surface area was incongruent with numerosity.

### 2.4.2 Results and Discussion

Overall, all three participants performed above chance at Time 1 (Feinstein 86%, $p < 0.001$; Mikulski 65%, $p < 0.001$; Schroeder 73%, $p < 0.001$) and Time 2 (Feinstein 88%, $p < 0.001$; Mikulski 68%, $p < 0.001$; Schroeder 77%, $p < 0.001$). All three monkeys also showed a reduction in WPre from Time 1 to Time 2 demonstrating an effect of training.

---

3 Schroeder was rewarded for ascending trials only at Time 2, but was differentially rewarded for ascending and descending trials, dependent on a color cue at Time 1. Only the data from ascending trials were analyzed.
(Figure 4. Feinstein 0.33 vs. 0.26; Mikulski 1.17 vs. 1.02; and Schroeder 0.66 vs. 0.54). A Bayesian analysis showed strong evidence for a training effect, indicating that $w_{Pica}$ values improved from Time 1 to Time 2. Feinstein and Schroeder both showed weber fractions roughly similar to those reported in Experiment 2 (Feinstein: Exp. 2 $w_{Pica} = 0.25$ vs. Exp. 3, Time 2 $w_{Pica} = 0.26$; Schroeder: Exp. 2 $w_{Pica} = 0.76$ vs. Exp. 3, Time 2 $w_{Pica} = 0.54$), indicating some continuity between performance on two different ordinal tasks. In contrast, Mikulski performed much worse on the ascending task than she did on the descending task (Exp. 2 $w_{Pica} = 0.38$ vs. Exp. 3, Time 2 $w_{Pica} = 1.02$).

![Graph showing decrease in weber fraction](image)

**Figure 4:** All three subjects in Experiment 3 showed a decrease in weber fraction ($w_{Pica}$) between Time 1 and Time 2.
2.4.2.1 Surface area

At each time point we conducted separate analyses on trials when 1) cumulative surface area was incongruent with numerosity, and 2) surface area was congruent with numerosity. We found that accuracy was above chance regardless of whether cumulative surface area was incongruent with numerosity (Time 1: M = 78.94%, t(2) = 10.80, p < 0.01; Time 2: M = 80.71%, t(2) = 13.44, p < 0.001) or cumulative surface area was congruent with numerosity (Time 1: M = 72.34%, t(2) = 11.59, p < 0.01; Time 2: M = 76.38%, t(2) = 11.91, p < 0.001).

We calculated a surface area difference score for each participant ($w_{Pca}$ for trials in which cumulative surface area was incongruent with numerosity subtracted from $w_{Pca}$ for trials in which surface area was congruent with numerosity). In contrast to the finding of both Experiment 1A which showed no consistent effect of surface area, and Experiment 1B which showed lower Weber fractions when surface area was congruent with numerosity, here we found positive difference scores for all subjects, indicating that Weber fractions were lower when cumulative surface area was incongruent with numerosity compared to when surface area was congruent with numerosity. This was true both at Time 1 (Difference scores: 0.28, 0.84, and 0.04 for Feinstein, Mikulski, and Schroeder, respectively) and at Time 2 (Difference scores: 0.24, 0.62, and 0.01). The decrease in difference scores from Time 1 to Time 2 are suggestive of the possibility that training decreases the effect of surface area congruence on numerical discriminations,
however the sample size is insufficient to determine if this minor effect is statistically significant.

2.5 General Discussion

Our first finding was that prosimian primates and rhesus monkeys showed quantitative equivalence in numerical discrimination despite approximately 50 million years of evolutionary divergence and significant differences in their visual system and visual acuity⁴. Thus our results suggest that previous findings of inferior numerical discrimination in lemurs relative to macaques were due to differences in training or task parameters and did not reflect ability differences between macaques and lemurs (Merritt, et al., 2011). These findings highlight the need to conduct comparative experiments with common methods and conditions.

These findings should also be interpreted in light of a recent study that suggests that there are no qualitative differences in the numerical discrimination abilities of rhesus macaques and pigeons (Scarf, et al., 2011). In that study pigeons successfully transferred an ordinal numerical rule from the values 1-4 to the values 5-9 much like rhesus monkeys (Brannon & Terrace, 1998). Pigeons showed ratio dependent

⁴ These results are especially surprising given the differences in the visual systems of lemurs and monkeys. Macaques, like apes and humans, are trichromats. The lemurs used in this experiment have photopigments consistent with dichromacy (Jacobs & Deegan, 1993, 2003; Leonhardt, Tung, Camden, Leal & Drea, 2008), although there is behavioral evidence for color discrimination consistent with trichromacy in ring-tailed lemurs (Blakeslee & Jacobs, 1985) and black lemurs (Gossett & Roeder, 2000). Lemurs have also been shown to have lower visual acuity than old world monkeys (Veilleux & Kirk, 2009). Despite these differences, we found no numerical acuity differences between lemurs and macaques using visual stimuli. This is likely due to the high contrast used in our visual stimuli.
performance and although the authors did not model weber fractions, performance looked very similar across these two species despite over 200 million years of divergent evolutionary history.

In addition to directly comparing the numerical abilities of lemurs and monkeys, a second goal for the present study was to compare numerical acuity in three different lemur species to assess whether numerical acuity might have been subject to selection pressures exerted by variations in social structure or feeding ecology. Our results suggest that despite the large variation in social structure, home range size, and diet in these three lemur species there is much more variability within a species than between species. The lack of species differences here is striking in light of other studies that compared closely related species and found differences that were consistent with evolutionary predictions (Bond, Kamil & Balda, 2003; MacLean, Merritt & Brannon, 2008; Platt, Brannon, Briese & French, 1996; Rosati, Stevens, Hare & Hauser, 2007; Tomasello, Hare & Fogleman, 2001). In prior research, ring-tailed lemurs have often exceeded the performance of other lemur species and performed more similarly to macaques than other lemurs. For example, Sandel, MacLean and Hare (2011) showed that ring-tailed lemurs, but not mongoose lemurs, black lemurs or ruffed lemurs, spontaneously exploited social cues regarding a competitor’s visual orientation. These similarities may reflect cognitive adaptations for living in large social groups rather than a common ancestral trait. Similarly, MacLean, et al. (2008) demonstrated that ring-tailed
lemurs show advanced transitive reasoning skills relative to mongoose lemurs, consistent with the prediction that animals that live in complex social structures have evolved cognitive adaptations for tracking dominance relationships.

In contrast, MacLean, et al. (2012) recently reported that differences in group size in 6 prosimian species predicted performance on a measure of social cognition, but not a nonsocial inhibitory control task. Our finding of a lack of species differences despite the large variation in social structure in a nonsocial cognitive measure of numerical cognition provides convergent evidence against the hypothesis that living in large social networks was the primary selective pressure for the evolution of nonsocial cognition. It must be emphasized, however, that power analyses indicated that our sample sizes may not have been large enough to detect species differences in weber fraction, were they to exist, due to large within-species variability. Obviously, samples of 3200 or even 36 individuals per species are not feasible for comparative cognitive research with prosimian primates. Thus, a caution is that small sample sizes with comparative research will only be useful insofar as they will uncover relatively large species differences that surpass individual variability within a species.

All subjects showed ratio-dependent performance characteristic of the approximate number system, and performed above chance regardless of surface area congruence, indicating that they were attending to numerosity rather than continuous properties of the stimuli. These results provide convergent evidence that humans,
monkeys and lemurs all share a system for representing number approximately. Many species from a wide variety of taxa as distantly related as fish (e.g. Agrillo, et al., 2008) and rats (e.g. Meck & Church, 1983) show ratio-dependent numerical discriminations. However, although many different animal species have been tested in a wide variety of tasks, rarely have multiple species been tested with the same task and stimuli to allow for a quantiative comparison. Rather than simply failing to reject the null hypothesis that the four species did not differ in numerical acuity, our Bayesian analytic approach provides strong evidence against pronounced species differences. We compared the strength of two different models: 1) the weber fractions of all subjects, regardless of species, were drawn from the same distribution and 2) that there are species differences in weber fraction for numerical acuity. Our analyses revealed strong support for models that assume a single distribution of \( w \)’s across all nonhuman species compared to models in which each non-human primate species was assumed to have a unique distribution of \( w \)’s.

A third central question addressed by our research was whether the ANS can be honed with extensive practice. Experiment 2 suggested that monkeys with extensive numerical experience had lower weber fractions than the monkeys who began Experiment 1A without any number training. The same Bayesian tests used to demonstrate a lack of species differences revealed that the number-experienced macaques tested in Experiment 2 and the macaques with limited numerical experience
tested in Experiment 1A were drawn from distinct distributions of $w$’s. Experiment 3 supported this conclusion in that all three number experienced monkeys showed a reduction in weber fraction over 3-6 years of intervening number training.

Although practice improves ANS acuity, another notable aspect of our results is that individual variability in numerical acuity is marked in nonhuman primates and this variance is so large that it can overshadow practice effects. These findings are consistent with the large variability in $w$ seen in adult human samples (see Piazza & Izard, 2009 for a review). For example, $w$ ranges in human samples have been as wide as 0.18 to 0.76 (DeWind & Brannon, 2012) or even 0.22 to 1.50 (Gilmore, et al., 2011). Our own study (Experiment 1B), using limited stimulus exposure but otherwise identical parameters as that used to test lemurs and macaques in Experiments 1 & 2, also revealed individual variability with a range of 0.16 to 0.33. Future studies should explore whether this individual variation reflects differences in general cognitive ability or more specialized quantitative abilities by correlating performance across a wide range of tasks.

In some sense the lack of species differences coupled with the large amount of individual variability poses a conundrum. On the one hand, selective pressures for the existence of the ANS seem to have been so extreme and ubiquitous that there is little variation across species. On the other hand, there is considerable variability across individuals within each species. If the ANS is a cognitive mechanism evolved under strong selective pressure and closely related to fitness, this variability may be surprising.
Comparisons of heritability indicate that traits under stronger directional selection (e.g., fitness-related traits) have reduced genetic variance compared with traits with more distant connections to fitness (e.g. Fisher, 1930; Mousseau & Roff, 1987; Roff & Mousseau, 1987). However, an alternative hypothesis that may resolve this discrepancy is that traits closely related to fitness are expected to have higher additive genetic and nongenetic variability due to the number of genetic and environmental events that effect complex, behavioral traits (e.g. Houle, 1991, 1992). The considerable variability of acuity in the ANS within species, along with the improvement across training, may suggest that environmental factors and experience within the lifespan play an important role in numerical sensitivity.

In conclusion, while research over the last century has revealed numerical abilities in a wide range of animal species, rarely has a single task been used to allow quantitative comparisons between species. By employing a common set of task parameters and stimuli our experiments revealed that there is a surprising amount of overlap in numerical acuity between nonhuman primate species. While some aspects of nonhuman primate cognition appear to have been shaped by the socioecological milieu in which the species has evolved the approximate number system appears to be insensitive to these variations.

There is abundant evidence that approximate numerical representations are universal in human societies, appear early in human development and are shared by nonhuman animals (see Brannon, 2005, 2006; Feigenson, Dehaene, & Spelke, 2004; Nieder & Miller, 2004 for reviews). In all of these groups, approximate number discrimination appears to be governed by Weber’s law such that accuracy decreases and reaction time increases for numerical comparisons as the ratio between the two values (min/max) approaches 1. For adult humans, Weber’s law holds across many numerical formats such as spoken words, Arabic numerals, tones, and dot arrays (Dehaene & Akhavein, 1995; Pica, et al., 2004; Whalen, et al., 1999; Dehaene, et al., 2008).

A lesser known characteristic of adult human comparisons is the semantic congruity effect. When asked to compare two large items, adults are quicker to respond to the question “Which is larger?” than “Which is smaller?” Conversely, when asked to compare two small items, adults are quicker to respond to the question “Which is smaller?” than “Which is larger?” The semantic congruity effect has been demonstrated with a variety of stimuli, including animal sizes, distance between cities, line length, brightness, surface area, Arabic numerals, and the numerosity of dot arrays (Audley & Wallis, 1964; Banks, et al., 1976; Čech & Shoben, 1985; Čech, et al., 1990; Holyoak, 1978; Holyoak & Mah, 1982; Moyer & Bayer, 1976; Petrusic, et al., 1998). For example, when
evaluating Arabic numerals, adults indicate that 2 is smaller than 3 more quickly than they indicate that 3 is larger than 2. However, when presented with two relatively large values such as 8 and 9, adults are faster when asked to indicate the larger compared to the smaller value (Banks, et al., 1976).

One proposal that accounts for the semantic congruity effect is the propositionally based semantic theory also known as the Discrete Code model (Banks, 1977; Banks, Clark, & Lucy, 1975; Banks, et al., 1976; Čech, et al., 1990.) According to this model, when two stimuli are compared on the basis of size or number, they are coded for magnitude (Small or Large) using linguistic processes. When the stimuli being compared differ greatly in magnitude, they are given different magnitude codes (e.g. Small vs. Large). If asked “Which is larger?” or “Which is smaller?” these codes can then be quickly and easily matched to the form of the question. However, when the stimuli being compared are either both large or both small, then the magnitude codes for both stimuli will be similar to one another (e.g. Large and Large+ or Small and Small+, respectively). If both stimuli have similar magnitude codes, and the form of the question does not match those codes, (i.e. when both stimuli are small, but the instruction is “choose larger”), then the codes must be translated from Small/Small+ to Small/Large so that one of the codes is consistent with the form of instruction. Under this proposal, the reaction time patterns that characterize the semantic congruity effect are the result of the added processing time necessary to translate these codes.
In contrast to the idea that semantic congruity effects emerge from linguistic coding, Cantlon and Brannon (2005) demonstrated a numerical semantic congruity effect with a nonverbal species: the macaque monkey (*Macaca mulatta*). Monkeys were trained on a touch screen computer to choose between two arrays of dots. A color cue indicated whether they would be rewarded for selecting the array with the larger (blue background) or smaller (red background) number of dots. The monkeys were tested with pairs of dots that ranged in value from 1–9. Monkeys showed a semantic congruity effect similar to that reported for humans: with a red background, they were faster to respond that 2 was smaller than 3 than they were to respond that 3 was larger than 2. In contrast, when the values were relatively large (i.e. 7 vs. 8), monkeys were faster to indicate the larger of the two values than they were to indicate the smaller of the two values with a blue background. Evidence of the semantic congruity effect in a non-linguistic animal species demonstrates that the effect must not depend on linguistic coding.

An alternative model, proposed by Holyoak and colleagues (Holyoak, 1978; Holyoak & Mah, 1982), does not rely on linguistic processes and instead assumes that humans compare stimuli with respect to a reference point which is either established implicitly by the form of the question or is presented explicitly by the experimenters. Holyoak (1978) tested his reference point model by presenting participants with an explicit reference point on a numerical continuum and asking them to determine which
of two digits in the range 1–9 was closer to that reference point. When the reference point was set at 1 or 9, reaction times showed a clear congruity effect similar to that reported by Banks, et al. (1976). When the reference point was set at a value in the middle of the range of values presented, reaction time was fastest for the pairs closest to the reference point provided.

The reference point model suggests that the endpoints of the continuum being tested serve as temporary reference points, and therefore, reaction time patterns for a given pair are dependent on whether the pair is considered to be relatively small or large within the range of values being tested. Thus, if a given pair of numerosities (or digits) were close to the higher endpoint of the testing range, then participants would be faster to choose the larger value in the pair (e.g., 8 vs. 9 in the range 1–9). If the range of values were then changed such that the same pair was closer to the lower endpoints, we would expect the subjects to be faster at indicating the smaller value of the pair (e.g., 8 vs. 9 in the range 8–64).

Čech and Shoben (1985) demonstrated this context effect for semantic congruity in adult humans. In their study, participants were presented with pairs of animals and asked which was larger or which was smaller. When only small animals were presented, participants were faster at choosing the larger of the two largest animals in that range (i.e. Rabbit–beaver), but when those same animals were presented in a broader range, the smaller of the two was identified more rapidly. Thus, the speed of human
comparison judgments is determined in part by the range of choice options available (i.e., the context of the judgment).

Like the semantic congruity effect, the context effect is a signature of the mental comparison process that has been shown in human adults. If monkeys and humans are using a homologous mental comparison process, then the range effects shown by humans should also be evident in the reaction time patterns shown by monkeys making numerical comparisons. The goal of the current paper is to examine the role of context on the semantic congruity effect in the numerical judgments of nonhuman primates. If context affects numerical semantic congruity in nonhuman primates, then the effect cannot be attributed to uniquely human processes such as linguistic encoding. Instead, it would provide evidence that despite the capacity for linguistic coding in humans, humans likely use a non-linguistic comparison process that results in a semantic congruity effect.

3.1 Experiment 1

3.1.1 Methods

3.1.1.1 Subjects and apparatus

Subjects were 2 socially-housed adult female rhesus macaques (*Macaca mulatta*), Mikulski and Schroeder. Both monkeys were kept on a water-restricted diet approved by an institutional animal care and use committee. Subjects were tested in sound-
attenuated booths while seated in Plexiglas primate chairs fitted with a water delivery system. Stimuli were presented on a touch screen computer monitor.

3.1.1.2 Task and Procedure

Monkeys were tested in an ordering task with both a small and large range of numerical stimuli. The two ranges were presented in blocks of 5 sessions\(^1\) in an ABBA design where A refers to the small range and B refers to the large range. Each session contained 200 trials. Each pair was presented with approximately equal frequency either 33 or 34 times within each session (randomized).

Subjects initiated each trial by touching a start stimulus, a small red square in the bottom-right corner of the screen. Pressing the start stimulus caused it to disappear, after which, two arrays of dots, each of which contained between 1 and 72 elements, were presented on the left and right sides of the screen (Figure 5).

\(^{1}\) The first A block for Schroeder contained only 4 sessions due to an experimenter error. The data for Schroeder’s missing session (fifth session in the first block of the smaller range) was replaced by the average accuracy from the first four sessions within the small range.
Figure 5: (A) Example of an ascending trial in small range, (B) example of a descending trial in small range, (C) example of an ascending trial in large range, (D) example of a descending trial in large range. Note that the background screen color was red for all ascending trials and blue for all descending trials.

If the background color of the screen was red, subjects were rewarded for selecting the array with the smaller number of elements first (“ascending” trials). If the background was blue, reward was given for selecting the array with the larger number of elements first (“descending” trials). Ascending and descending trials were randomly intermixed within a session. A trial was terminated after the arrays were touched in the correct order, or when the subject made an incorrect response. Correct trials were rewarded with 1.16 ml of water and positive auditory feedback, which was followed by
a 1 s intertrial interval (ITI). Incorrect responses ended the trial and initiated a 2 s timeout added to the intertrial interval (ITI) and negative auditory feedback.

3.1.1.3 Prior Training

Both monkeys had extensive prior numerical training, including numerical matching and ordinal tasks. Mikulski had experience ordering all pairs of 1–9, as well as larger numerosities up to 150. Schroeder had more limited prior training ordering pairs with a 1:2 ratio with absolute values between 1 and 150. Prior training for Mikulski is described in Cantlon and Brannon (2005) and training for Schroeder was similar.

3.1.1.4 Testing

Subjects were tested using the task described above for two ranges. The small range contained the following numerosity pairs: 1:2, 2:4, 3:6, 4:8, 5:10, 6:12. The large range was established by multiplying the small range by 6, thereby creating the pairs 6:12, 12:24, 18:36, 24:48, 30:60, 36:72.

3.1.1.5 Stimuli

Stimuli were randomly generated by a RealBasic program and were thus trial unique according to parameters that allowed surface area, element size and color to be controlled. All elements within a stimulus were square and the background of the array was always yellow. Individual elements were homogeneous in color within an array and the color was randomly selected for each stimulus from a list of 16 colors (brown, red, green, turquoise, pink, dark blue, orange, gray, blue, dark green, periwinkle, purple,
teal, dark red, dark purple and dark pink). Elements within each stimulus were also homogeneous in size and varied between 0.04 and 36 cm$^2$.

To control for cumulative surface area and individual element size, surface area was equated for the two arrays on half of the trials, whereas element size was equated for the two arrays on the other half of the trials. To control for density, the density of the two arrays was congruent with number on half of the trials and was either equated for the two arrays or incongruent with number on the other half of the trials.

### 3.1.2 Results

#### 3.1.2.1 Accuracy

Overall, monkeys performed significantly above chance across all test sessions on both the ascending trials [single-sample $t$ tests vs. 50% (chance), Mikulski, mean accuracy 73%, $t(19) = 13.01, p < 0.001$; Schroeder, mean accuracy 66%, $t(19) = 11.63, p < 0.001$] and the descending trials [single-sample $t$ tests: Mikulski, mean accuracy 70%, $t(19) = 15.48, p < 0.001$; Schroeder, mean accuracy 63%, $t(19) = 10.54, p < 0.001$]. Performance did not differ between ascending and descending trials [Mikulski, $t(19) = 0.74, p = 0.47$; Schroeder, $t(19) = 1.47, p = 0.16$].

Monkeys performed significantly above chance both when density was congruent with number [single-sample $t$ tests: Mikulski, $t(19) = 27.72, p < 0.001$;
Schroeder, $t(19) = 16.91$, $p < 0.001$] and when density was not congruent with number\(^2\) [single sample $t$ tests: Mikulski, $t(19) = 20.69$, $p < 0.001$; Schroeder, $t(19) = 9.75$, $p < 0.001$].Similarly, monkeys performed significantly above chance both when cumulative surface area was congruent with number [single-sample $t$ tests; Mikulski, $t(19) = 27.24$, $p < 0.001$; Schroeder, $t(19) = 20.84$, $p < 0.001$] and when cumulative surface area was equated between the two arrays [single-sample $t$ tests: Mikulski, $t(19) = 15.71$, $p < 0.001$; Schroeder, $t(19) = 6.85$, $p < 0.001$]. Thus, both monkeys used numerical value, rather than density or cumulative surface area, to compare stimuli.

3.1.2.2 Semantic Congruity Effect

The main finding was that both monkeys showed a semantic congruity effect for the small and large numerosity ranges (Figure 6). In the small range, an ANOVA for Direction (ascending, descending) $\times$ Numerical Magnitude [Small (1 vs. 2, 2 vs. 4, 3 vs. 6), Large (4 vs. 8, 5 vs. 10, 6 vs. 12)] on RT revealed an interaction between the Direction and the Numerical Magnitude of the comparison stimuli for each monkey [Mikulski, $F(3,39) = 29.34$, $p < 0.001$; Schroeder, $F(3,39) = 38.45$, $p < 0.001$]. In addition there was a main effect of Direction for both monkeys [Mikulski, $F(3,39) = 80.64$, $p < 0.001$; Schroeder, $F(3,39) = 14.88$, $p < 0.001$], reflecting overall faster RTs on ascending trials than descending trials [Mikulski, $t(19) = 6.07$, $p < 0.001$; Schroeder, $t(19) = 2.38$, $p < 0.05$].

\(^2\)Trials in which density was equated between the two arrays (30% of trials) and trials in which density was incongruent with number (20% of trials) were collapsed for analysis.
Finally there was a main effect of Numerical Magnitude for Schroeder \(F(3,39) = 6.42, p < 0.05\), reflecting moderately faster RTs on large pairs compared to small pairs in the small range \(t(19) = 1.54, p = 0.14\). Note, that for this and all other analyses, only RTs to correct responses were included.

![Figure 6](image)

**Figure 6:** (A) Mean RTs for Mikulski in small range, (B) Mean RTs for Mikulski in large range, (C) Mean RTs for Schroeder in small range, (D) Mean RTs for Schroeder in large range. Both monkeys showed a semantic congruity effect in both ranges.

In the large numerosity range, an ANOVA for Direction (ascending, descending) \(\times\) Numerical Magnitude [Small (6 vs. 12, 12 vs. 24, 18 vs. 36), Large (24 vs. 48, 30 vs. 60, 36 vs. 72)] on RT also revealed an interaction between the Direction and the Numerical...
Magnitude of the comparison stimuli for both monkeys [Mikulski, $F(3,39) = 21.11, p < 0.001$; Schroeder, $F(3,39) = 20.56, p < 0.001$]. Thus, both monkeys showed a semantic congruity effect for both the small and large numerical ranges.

In addition to the interaction between Direction and Numerical Magnitude in the large range there was a marginally significant main effect of Direction for Schroeder [$F(3,39) = 3.80, p = 0.06$], indicating overall faster RTs on descending than ascending trials. Finally there was a main effect of Numerical Magnitude for Mikulski [$F(3,39) = 17.85, p < 0.001$], reflecting overall faster RTs on small pairs than large pairs in this range [$t(19) = 3.20, p < 0.01$].

### 3.1.2.3 Context Effect

As can be seen in Figure 6, Mikulski’s semantic congruity effect for the small range showed a “funnel effect” whereby RT was significantly shorter for small values on ascending trials compared to descending trials and there was no difference between the two trial types for large values. In contrast for Schroeder on both ranges and for Mikulski on the large range, the semantic congruity effect was a true crossover effect (Holyoak, 1978). The finding that the semantic congruity effect held for both ranges indicates that the monkeys treated the largest values in the small range as small values when they were presented within the context of the larger range. To more closely examine this context effect we examined RT on the 6:12 pair which was a critical pair because it was the largest pair of values in the small range and the smallest pair of
values in the large range. As shown in Figure 7, monkeys were faster on average on ascending trials than descending trials when 6:12 was a relatively small pair (in the large range). In contrast, monkeys were slightly faster on average on descending trials than on ascending trials when 6:12 was a relatively large pair (in the small range).

![Bar chart](image)

**Figure 7:** Mean RTs averaged across both monkeys on the pair 6 vs. 12 for all correct trials in the small and large ranges. When the pair 6 vs. 12 was the largest set of values in the small range, monkeys were, on average, faster to choose the larger than to choose the smaller of the two arrays. In contrast, when the pair 6 vs. 12 was the smallest set of values in the large range, monkeys were, on average, faster to choose the smaller than to choose the larger of the same two arrays.

An ANOVA for Direction (ascending, descending) × Range [Small (1:2 through 6:12), Large (6:12 through 36:72)] on Schroeder’s RT to the pair 6:12 revealed a main effect of Range [$F(3,39) = 11.76, p < 0.01$], reflecting her overall faster responses on trials in the small range than trials in the large range, as well as a significant interaction between the Range and the Direction in which she was required to respond (ascending vs. descending) [$F(3,39) = 17.67, p < 0.001$]. This interaction indicates that, when the pair
6:12 was a relatively large pair in the small range, Schroeder was faster to choose the larger than to choose the smaller of the two arrays \( t(9) = 3.20, p < 0.05 \), whereas when the same numerosity pair was presented in the large range and therefore was a relatively small pair, Schroeder was faster to choose the smaller than to choose the larger of the two arrays \( t(9) = 3.76, p < 0.01 \).

The same analysis on Mikulski’s RT revealed a main effect of Direction \( F(3,39) = 4.76, p < 0.05 \), reflecting her overall faster responses to the pair 6:12 on ascending trials than descending trials. She also showed a trend towards a significant interaction between the Range and the Direction of the comparison stimuli \( F(3,39) = 2.34, p = 0.13 \). Thus although Mikulski showed a trend towards the predicted pattern, only Schroeder showed a significant contextual shift.

### 3.1.2.4 Timing of contextual shift

A final analysis assessed when the contextual shift took place by examining response times to the smallest pairs (6 vs. 12 and 12 vs. 24) in the large range. A contextual shift would result in monkeys being faster at ascending trials for the pair 6 vs. 12 in the large range, despite the fact that it was previously the largest pair in the small range (Figure 8). Given how few trials were available for the pair 6 vs. 12 from each session, we included the next smallest pair, 12 vs. 24, to increase power. A contextual shift would also result in monkeys being faster on ascending trials for 12 vs. 24, even though these values were larger than any pair in the small range. We conducted t-tests
comparing the RTs for these pairs on ascending vs. descending trials on the sessions following the switch from the small to the large range. Despite the fact that both monkeys, at the end of small range testing, were faster at ordering the largest pair (6 vs. 12) when given a descending compared to ascending cue, both monkeys quickly became faster at ordering the 6 vs. 12 and 12 vs. 24 pairs in ascending order when these pairs were presented in the large range context. Schroeder was faster on ascending than descending trials on these pairs by the third session of large range testing [$t(47) = -2.020$, $p < 0.05$] and Mikulski showed the same effect on the very first large range session [$t(44) = -2.202$, $p < 0.05$].

![Figure 8: Median RTs for descending trials for the pair 6 vs. 12 subtracted from median RTs for ascending trials for the pair 6 vs.12. Data reflects the average of the median RTs for correct trials for the 5 sessions of each block averaged for the two monkeys. Positive values indicate that the monkeys were faster on descending trials compared to ascending trials for the pair 6 vs. 12, whereas negative values indicate the reverse. Monkeys showed a rapid contextual shift between blocks such that in blocks of small range sessions, monkeys were faster on descending trials, and in blocks of large range sessions, monkeys were faster on ascending trials for the 6 vs. 12 pair.](image-url)
3.1.3 Discussion

Our findings provide further evidence against linguistic coding models of the semantic congruity effect. Both monkeys showed striking semantic congruity effects replicating a prior report by Cantlon and Brannon (2005). The novel aspect of our study, however, is that the semantic congruity effect in monkeys is influenced by context as was previously reported in humans (Čech and Shoben, 1985). Specifically, for at least one monkey, a given pair of values was treated as small or large depending on the range of values in which it was presented. Thus our data indicate that the cognitive process that monkeys use to compare the numerical value of two arrays of dots is sensitive to the range of numerosities being presented and thus shares yet another important feature with the comparison processes of human adults: the context effect of numerical comparisons.

Another question that could be addressed by our results was how quickly the monkeys responded to a shift in context. Čech, et al. (1990) and Petrusic and Baranski (1989) found that adult humans respond to a shift in context very rapidly (e.g., within 40–92 trials). Čech, et al. (1990) asked human adults to select either the larger or the smaller of a pair of animals. Small animals were paired only with other small animals and large animals were paired only with other large animals (the ranges did not overlap: all small animals were smaller than all large animals). The authors reported a context effect in RTs to the largest pair in the small range and the smallest pair in the large range.
that was established within the first 40–92 trials. The authors suggest that initial sampling of pairs may determine how subjects determine what is relatively large or small.

Petrusic and Baranski (1989) similarly reported that adults rapidly adjust the way they code stimuli based on context. Pairs of weighted objects were presented in blocks of 46 trials and participants were asked to choose either the lighter or the heavier object. Data from the first block of trials showed a funnel effect, while data from the second block of trials showed a full crossover semantic congruity effect. The authors interpreted this to indicate that the reference point against which the stimuli were compared shifted rapidly (within the first 92 trials) to one more appropriate for the range of the stimuli. Our analyses indicate that like adult humans, monkeys are capable of rapidly shifting their response strategies and showing a contextual shift within the first few sessions in which the range of values is changed.

In summary, our findings indicate that the numerical comparison process of monkeys is sensitive to the range of numerosities being presented and provide further support for a numerical comparison process shared by monkeys and humans. The numerical semantic congruity effect is found in both species and is similarly sensitive to contextual anchoring. Monkeys respond to identical stimuli differently depending on the context in which those stimuli are presented. We suggest that an adequate model of
the comparison process must account for both the semantic congruity effect and the context effect without invoking linguistic coding.

Preverbal human infants and a few nonhuman animal species have exhibited two contrasting patterns of behavior when faced with quantity judgments. In some tasks, performance is independent of set size and is modulated by the ratio between the two values being compared. In other tasks, successful discrimination is limited to very small values and shows no signs of ratio dependence. This has led to the proposal that there are two cognitive systems that underlie nonverbal numerical discrimination: a limited capacity object file system, which allows the accurate representation of a small number of objects through attentional tracking, and an approximate number system (ANS), which is ratio-dependent and has no upper limit in its capacity (Feigenson, et al., 2004; Uller, Carey, Huntley-Fenner, & Klatt, 1999). The ANS is ubiquitous throughout the animal kingdom and has been shown to operate for large values throughout human development and adulthood (for reviews, see Beran, 2008b; Brannon, 2006). The object file system has been well-documented in human infants under a limited set of circumstances (e.g. Feigenson & Carey, 2003, 2005; Feigenson, et al., 2002), and to a much lesser extent, in non-human animals (e.g. Agrillo, et al., 2007, 2008; Hauser, et al., 2000; Rugani, et al., 2008; Uller & Lewis, 2009).

Object files are not explicitly numerical representations, but instead represent individual objects in attention. Each object file “sticks” to a unique object as it moves
about the visual scene, and may contain identity or featural information (Kahneman, et al., 1992). The object file system represents individuated objects, with the number of open object files providing an implicit way to represent the numerosity of a set. However, as only 3 or 4 object files can be maintained simultaneously, the ability of this system to provide a means of representing numerosity is limited to small numbers.

In contrast, the ANS represents the cardinality of a set of objects as a single mental magnitude. The ability to discriminate between two numerosities in the ANS is ratio-dependent, in accordance with Weber’s Law and is not limited by set size. Small values that are within the capacity of the object file system could therefore be represented with greater precision as object files given the ratio-dependent property of the ANS. Thus babies and animals, both of which lack a verbal counting system, could potentially maximize reward in food choice paradigms were they to use the object file system to discriminate small pairs accurately and the ANS to discriminate large pairs approximately. This pattern of responding is uncommon. Agrillo, et al. (2012) reported that for both adult humans and guppies, accuracy at selecting the larger numerical value was affected by ratio for large values, but not for the numbers 1–4. Likewise, Agrillo, et al. (2008) reported that mosquitofish discriminated between small sets 1 vs. 2, 2 vs. 3 and 3 vs. 4, were able to discriminate large values only when those values differed by a ratio of 1:2 or easier (4 vs. 8, 8 vs. 16 and 4 vs. 10).
More common in the numerical discrimination literature are studies in animals and human infants either show ratio dependence across the entire range (e.g., Cantlon & Brannon, 2006b; Beran, 2007; vanMarle & Wynn, 2009) or they show a set size limit such that if either numerosity exceeds the limit, discrimination drops to chance levels of accuracy (e.g., Feigenson, et al., 2002; Hauser, et al., 2000). There is convergent evidence from multiple behavioral paradigms that human infants discriminate between small numerosities (<3) accurately (e.g., Feigenson, et al., 2002; Feigenson & Carey, 2003; Koechlin, Dehaene, & Mehler, , 1997; Starkey & Cooper, 1980; Strauss & Curtis, 1981; Wood & Spelke, 2005; Wynn, 1992; Xu, 2003). A subset of these studies have provided strong evidence for the object file system as opposed to the ANS, specifically success with small values (< 3) at a given ratio and simultaneous failure with large sets (>3) at the same ratio (e.g., 2 vs. 3 and 6 vs. 9). The food choice task used by Feigenson and colleagues has repeatedly shown a set-size limitation in quantity discriminations in infants (Feigenson & Carey, 2005; Feigenson, et al., 2002.). In this paradigm, infants are shown food items being dropped into two opaque containers and then allowed to approach one of the containers and consume its contents. Feigenson, et al. (2002) demonstrated that 10 to 12 month old infants reliably crawled to the container with a greater number of food items when both contained 3 or fewer food items. Thus, 10 and 12 month old infants succeeded at choosing the larger in a 1 vs. 2 and a 2 vs. 3 condition, but performed at chance in a 2 vs. 4 or a 3 vs. 6 condition. Controls for overall duration,
complexity, and motivation caused no change to this pattern of performance. In a separate experiment, they demonstrated that infants performed at chance in a 1 vs. 4 condition, but successfully chose the larger in a 0 vs. 4 condition, indicating that infants were capable of representing the existence of crackers, even if they couldn’t represent the actual number (Feigenson & Carey, 2005).

Feigenson and Carey (2003, 2005) demonstrated the same set size limit in 10 to 14 month old infants using a manual search task. When 3 toys were placed in a box and the experimenter retrieved 2, the subject spent more time searching for the missing toy than when 2 toys were placed in the box and the experimenter retrieved both. This difference in search time indicates that the subject represented both the number of toys placed in the box and the number of toys removed, and noticed the discrepancy. In contrast, when 4 toys were placed in the box and 2 were retrieved, subjects did not search longer than when all items had been removed from the box, suggesting that they could not distinguish between 2 and 4 items. As in Feigenson, et al. (2002), infants were able to discriminate 1 from 2, and 2 from 3 items, but were unable to distinguish 2 from 4 items. Together, these findings from the food choice task and the manual search task indicate that infants may rely on a system that cannot handle more than three objects.

Set size limitations consistent with the object file system have also been reported in numerical discriminations by nonhuman animals, although there is far more evidence for the ratio-dependent hallmark of the ANS. Hauser, Carey and Hauser (2000) used a
single-trial food choice task similar to the cracker choice task used by Feigenson, et al. (2002) with semi free-ranging, untrained rhesus macaques. Monkeys watched as apple slices were placed into each of two opaque boxes. Monkeys were then allowed to approach and consume the apple slices in one box. The monkeys chose the greater number of apple slices as long as the contents of each box did not exceed the set size limit of 4. However, on many comparisons where one box exceeded that limit the monkeys performed at chance, showing no preference for the greater number of food items. This was true even with favorable ratios: 4 vs. 8 and 3 vs. 8. Oddly however, monkeys successfully discriminated 3 vs. 5 in the same study. Wood, Hauser, Glynn and Barner (2008) used a similar food choice task with non-solid food portions in the same population of rhesus macaques. The monkeys spontaneously chose the box containing the larger number of portions of carrot pieces, poured from a cup, for 1 vs. 2, 2 vs. 3, and 3 vs. 4 portions, but not 4 vs. 5 or 3 vs. 6 portions (despite the equal ratio of 3 vs. 6 and 1 vs. 2).

Beyond primates, set size limitations have been demonstrated in animals as diverse as horses, insects, amphibians and birds. Uller and Lewis (2009) performed an experiment similar to the one reported by Hauser, et al. (2000) demonstrating quantity discrimination in horses. Horses successfully chose the larger of 1 vs. 2 and 2 vs. 3 apples, but failed to systematically choose 6 over 4 apples. Fish and amphibians may also rely on object files to select between two small sets. Red-backed salamanders chose
the larger set when presented with 2 vs. 3 fruit flies, but not 4 vs. 6 fruit flies despite the ratio equivalence (Uller, et al., 2003). Mosquito fish were also found to show a set size limit in discriminating between shoals (Agrillo, et al., 2007). When in a new environment, female mosquito fish will join a shoal, or collection of other female fish, for protection. Agrillo, et al. (2007, 2008) found the set size limit (<4) characteristic of the object file system in numerical comparisons that differed by 1, such that fish were more likely to move towards the larger of two shoals in comparisons of 1 vs. 2, 2 vs. 3, and 3 vs. 4, but were not more likely to select the larger shoal for comparisons of 4 vs. 5, 5 vs. 6, 6 vs. 7, or 7 vs. 8. Importantly, Agrillo and colleagues also showed evidence of ratio-dependent performance with large numbers such that they were able to discriminate large numbers at a 1:2 ratio (e.g., 8 vs. 16), but failed at a 2:3 ratio (8 vs. 12).

Only one study has tested numerical abilities in infant non-human animals. Rugani, Regolin and Vallortigara (2008) trained young chicks to peck at arrays of dots depending on their numerosity. Chicks successfully learned to discriminate 1 vs. 2 and 2 vs. 3, but failed to learn to discriminate 3 vs. 4, 4 vs. 5, or 4 vs. 6. Given success with 2 vs. 3 and failure with 4 vs. 6, this suggests that the animals were using the object file system rather than the ANS.

One possibility is that untrained animals will attempt to use the object file system whenever they are faced with quantity comparisons. However, this hypothesis is not supported by the fact that untrained animals have exhibited ratio-dependent
performance indicative of the ANS. Specifically, Flombaum, et al. (2005) used a violation of expectancy paradigm to assess large number discrimination and spontaneous expectations about addition in semi-free ranging macaques. Untrained rhesus macaques watched a series of events in which a set of food items was presented and hidden by a screen, after which another set of food items was added. When the screen was lowered, rhesus monkeys looked longer at numerically impossible outcomes to the addition operation than at numerically possible outcomes. Despite their lack of training, the monkeys’ limit on this ability was set by the ratio difference between the two numerical values. Likewise, Hauser, Tsao, Garcia, and Spelke (2003) used a different measure of spontaneous numerical discrimination to demonstrate ratio dependence in nonhuman primates. They familiarized cotton-top tamarins with sequences of auditory stimuli comprised of a specific number of syllables. At test, a novel number of syllables were played. Discrimination of the change in numerosity was operationalized as more frequent orienting towards the speaker. Untrained tamarins were found to discriminate numerical values that differed by a 1:2 or a 2:3 ratio, but not a 4:5 ratio. Thus, lack of training is insufficient to cue the object file system over the ANS for numerical comparisons. Nor is the food choice task itself sufficient to reliably tap the object file system. Nonhuman primates trained and tested on a food choice task have shown ratio-dependent discrimination of simultaneously visible sets (e.g. Addessi, et al., 2008;
Anderson, et al., 2007; Hanus & Call, 2007) and of sequentially presented sets (e.g. Beran, Evans, Leighty, Harris, & Rice, 2008; Hanus & Call, 2007; vanMarle, et al., 2006).

The majority of research on numerical abilities in nonhuman primates has focused on a few representative species: rhesus macaques, capuchin monkeys, and chimpanzees (e.g. Addessi, et al., 2008; Beran 2007; 2008a,c; Beran, Evans, & Harris, 2008; Boysen & Berntson, 1989; Brannon & Terrace, 1998; 2000; Cantlon & Brannon, 2006a,b, 2007a,b; Hauser, et al., 2000; Tomonaga, 2008). Very few studies have examined numerical abilities in prosimian primates (e.g. Jones, et al., under review; Lewis, et al., 2005; Merritt, et al., 2011; Santos, Barnes, & Mahajan, 2005). Including prosimian primates in comparisons of primate cognition is likely to be important in attempting to identify cognitive profiles of the primate ancestral state. Prosimian primates have been hypothesized to be morphologically and behaviorally similar to the last common primate ancestor (Tattersall, 1982; Yoder, 2007). Thus, if prosimians primates share cognitive traits that are common among other primates, it is likely that these traits were present in the last common ancestor.

Using a looking time measure adapted from the infant literature, Santos, Barnes, and Mahajan (2005) found that lemurs were spontaneously able to quantify a small set of objects and form precise expectations about a simple 1+1 addition event. Lewis, Jaffe and Brannon (2005) modified the manual search task from Feigenson & Carey (2003) to test spontaneous numerical comparisons in lemurs. Untrained mongoose lemurs
watched as grapes were sequentially placed into a bucket with a false bottom. The experimenter surreptitiously hid some subset of the grapes in the false bottom. When the lemurs were allowed to retrieve the grapes from the bucket, they were predicted to search longer if they expected there to be more grapes than they had already retrieved from the bucket. Thus, longer search times indicate that participants noticed a numerical discrepancy between the number of grapes they saw baited and the number they retrieved. Lewis, et al. (2005) reported longer search times when the difference between the number of grapes baited and the number retrieved was a 1:2 ratio compared to search time when all the baited grapes had been retrieved (i.e., 4 in 2 out vs. 2 in 2 out). In contrast there was no difference in search time when the ratio between baited and retrieved was 2:3 or 3:4.

In contrast to the Santos, Barnes, and Mahajan (2005) and Lewis, et al. (2005) studies which examined spontaneous numerical discrimination, Merritt, et al. (2011) and Jones, et al. (under review) used a touch-screen task to measure numerical comparison abilities in lemurs and macaques. Both studies showed the ratio-dependent hallmark of the ANS in these numerical comparisons. Jones, et al. (under review) tested three different lemur species and macaques and found overlapping numerical acuity for the four species. Thus, to date, both spontaneous numerical comparisons and training have led to evidence for the ANS in lemurs. However, it’s important to note that while Lewis, et al. (2005) used a spontaneous measure of numerical discrimination, each subject
participated in a few trials. As repeated trials may increase the likelihood of cuing the ANS (vanMarle, et al., 2006), it is unclear whether prosimian primates will show a set size limit consistent with object file representations, or the ratio dependence of the ANS, in a single trial measure of spontaneous numerical comparisons.

Here we use a modified spontaneous food choice task based on Hauser, et al. (2000) and Feigenson, et al. (2002) to test the spontaneous quantity discriminations of prosimian primates. Testing conditions differed from Hauser, et al. (2000) in that the subjects in the present study were separated briefly from conspecifics during testing to reduce interruptions and competition. Subjects from Hauser, et al. (2000) were semi free ranging macaques on the island Cayo Santiago. The monkeys living on Cayo naturally divided into several social groups, have unrestricted access to food and water, and are habituated to human experimenters. As the macaques are not caged, it is impossible to completely isolate a single animal for testing. Experimenters make an effort to engage the attention of animals that appear to be temporarily separate from the rest of their social group, but cannot eradicate the possibility that a conspecific may interfere. In contrast, the subjects in the present study were located at the Duke Lemur Center. All subjects were housed singly or in small groups of 2 to 8, in separate enclosures. This setting allowed us to separate lemurs from their conspecifics for testing, which allowed subjects an opportunity to participate in the task without competition or interruption from other animals.
Our study differed from Hauser, et al. (2000) in task design as well as testing conditions. A limitation of prior studies supporting the two systems-view of numerical cognition in nonhuman animals is that the conditions were not ideal to test for ratio dependence. To this end, we chose a set of values that included 2 sets of absolute values for each of three conditions (small-small, small-large, and large-large). The two sets of absolute values were at a 1:2 and 1:3 ratio for each condition.

An object-file system would be implicated if: 1. Accuracy is significantly above chance levels of performance, only when both numerosities are smaller than the set size limit. 2. Accuracy drops to chance for pairs of numerosities that exceed the set size limit even when the ratio between them is successfully discriminated with smaller numbers (e.g. success at 2 vs. 3 and failure at 4 vs. 6). In contrast, the ANS would be implicated if: 1. Lemurs successfully discriminate pairs with large values, 2. Lemurs show ratio-dependent response functions, with accuracy dropping to chance as the ratio (larger/smaller) approaches 1. A third possibility is that lemurs use object-files to represent small values and ANS representations to handle large values and that they are unable to compare these representations and consequently perform at chance on small-large comparisons (Table 2).
Table 2: The object file system, the ANS, and the two system theory of numerical discrimination offer distinct predictions for performance on the conditions from Chapter 4, Experiment 1.

<table>
<thead>
<tr>
<th>1:3 ratio</th>
<th>Quantities</th>
<th>Object File System</th>
<th>ANS</th>
<th>Both systems - Incommensurate representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vs. 3</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔ (object file)</td>
</tr>
<tr>
<td>2 vs. 6</td>
<td>X</td>
<td>✔</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4 vs. 12</td>
<td>X</td>
<td>✔</td>
<td>✔</td>
<td>✔ (ANS)</td>
</tr>
</tbody>
</table>

We further modified the protocol used by Hauser, et al. (2000) to eliminate the possibility of a Clever Hans effect. In the majority of quantity discrimination research, the experimenters presenting food items have been aware of which container held the larger quantity. It is thus possible that subjects made selections based on unintentional social cues from the experimenters. We established a simple modification to the design, which allowed the experimenters presenting food items to be blind to the condition on any given trial. At the time of testing, each experimenter was given an index card that indicated the number of food items they were to drop into their bucket and they were unaware of the number being baited in the other bucket.

4.1 Experiment 1

4.1.1 Methods

4.1.1.1 Subjects

Subjects were 113 diurnal and cathemeral prosimian primates (61 females and 52 males; mean age 13.01 years, SD 9.57), housed at the Duke Lemur Center. Seven of
these 113 animals were tested on two conditions. An additional 27 trials were excluded due to subject’s failure to participate (N = 22) or experimenter error (N = 5). Of these, 13 subjects were tested again at a later date in a randomly assigned condition. Participants included individuals from five different genuses and 15 different species (Appendix A). Members of each genus were equally distributed among the six conditions, such that each condition contained 5 Lemurs, 8 Eulemurs, 3 Varecia, 3 Propithecus, and 1 Hapalemur. All animal procedures were conducted in accordance with a Duke University IACUC protocol.

4.1.1.2 Procedure

Set up

Each subject remained in its home enclosure for testing, but was temporarily restricted from access to cagemates. Each subject was assigned quasi-randomly to one condition. Each trial involved three experimenters. One experimenter operated the camera (E1) and the other two experimenters dropped food items into the buckets (E2 and E3). The numerical conditions were assigned before testing and were known only to E1. E1 gave E2 and E3 each an index card that indicated the number of food items they were to drop into their bucket, which side they were to stand on (left or right), and whether they were to bait the bucket first or second. Experimenters were blind to the number of food items the other experimenter was baiting.
On each trial E2 and E3 stood 2-3 feet apart, immediately outside of the subject’s enclosure, and each held a black bucket that was approximately 12” in diameter and 10” in height. E1 stood behind the other two experimenters. At the onset of each trial E1 said “start” at which point the two experimenters faced the cage and tipped their buckets on their sides to show the subjects that the buckets were empty.

Presentation

E2 and E3 held the buckets with both hands at chest level. E3 closed his/her eyes and remained motionless as E2 baited the bucket with raisins or nuts (depending on dietary restrictions of each species). Each food item was removed from the experimenter’s left breast pocket and held up for the subject to see. Once the experimenter was certain the subject had seen the food item he/she placed it in the bucket. This was repeated until E2 had presented all food items, at which point he/she said “done” and closed his/her eyes. E3 then opened his/her eyes and baited the bucket following the same procedure including stating “done” and closing his or her eyes.

After all food items were presented, E1 determined when the subject had moved to a location approximately equidistant from both buckets and/or averted their gaze from either bucket. E1 then said “buckets down,” at which point E2 and E3 opened their eyes, crouched down and simultaneously set their buckets on the ground against the exterior cage wall. E1 and E2 then stood up, turned 180 degrees, and walked to the
other side of the hallway. E1 also turned 180 degrees and watched the subject’s choice via the small finder on the camera.

Selection

A trial ended when E1 determined that the subject had made a choice by moving in front of one of the two buckets and orienting towards it (Figure 9), or when 3 minutes passed and no choice was made. All data was re-coded by an independent observer who was blind to the hypotheses of the study.

Figure 9: Subjects made a choice by approaching one of the two buckets.

4.1.2 Results

E1 and the independent observer agreed on which bucket had been chosen on 116 trials (96.67% agreement). For the four trials on which they disagreed, an additional
experimenter blind to the condition coded the video and the majority decision was included. Furthermore, the coding of these four trials does not change the reported pattern of results.

Overall subjects selected the larger quantity more often than predicted by chance (82 out of 120 trials, \( p < 0.001 \)). Binomial sign tests indicated that subjects chose the larger number of food items significantly more often than predicted by chance for 1 vs. 3 (16 out of 20 trials, \( p < 0.01 \), one-tailed), 2 vs. 6 (16 out of 20 trials, \( p < 0.01 \), one-tailed), and 4 vs. 12 (15 out of 20 trials, \( p < 0.05 \), one-tailed). In contrast, binomial sign tests indicated that subjects chose the larger number of food items no more often than predicted by chance for 1 vs. 2 (14 out of 20 trials, \( NS \)), 3 vs. 6 (10 out of 20 trials, \( NS \)), or 6 vs. 12 trials (11 out of 20 trials, \( NS \)). It should be noted that subjects showed a trend towards selecting the larger number for the 1 vs. 2 condition (\( p = 0.06 \)) (Figure 10).
Figure 10: Results support the predictions of the ANS. Subjects successfully chose the larger in a small vs. large discrimination for conditions in which values differed by a 1:3 ratio, but not a 1:2 ratio.

4.2 Experiment 2

Given the trend towards selecting the larger in the 1 vs. 2 Condition, the goal of Experiment 2 was to test lemurs with a 2 vs. 3 comparison which has an unfavorable ratio but may be within the capacity of the object-file system.
4.2.1 Methods

4.2.1.1 Subjects

Subjects were 20 diurnal and cathemeral prosimian primates (13 females and 7 males; mean age 11.16 years, SD 8.74), housed at the Duke Lemur Center. An additional 5 trials were excluded due to subject’s failure to participate. Subjects represented a similar distribution of species as reported for the conditions in Experiment 1. Due to a limited number of naïve animals available for testing, 4 out of the 20 subjects had already participated in Experiment 1.

4.2.1.2 Procedure

The procedure was identical to Experiment 1.

4.2.2 Results

Subjects selected the larger quantity no more often than predicted by chance for 2 vs. 3 (8 out of 20 trials, NS) (Figure 11). The 4 subjects who had been tested in Experiment 1 showed no consistent pattern of responding and the exclusion of these trials would not change the pattern of results: One chose the larger in both experiments, one chose the smaller in both experiments, one chose the larger in Experiment 1 but the smaller of 2v3, and one chose the smaller in Experiment 1 but the larger of 2v3.
Figure 11: Subjects chose the larger of the 1 vs. 2 (nonsignificant trend, $p = 0.06$) and 1 vs. 3 ($p < 0.01$) conditions in Experiment 1, but failed to choose the larger of 2 vs. 3 in Experiment 2, implicating the ANS rather than the object file system.

4.3 Experiment 3

The goal of Experiment 3 was to provide a scent control. To this end we tested lemurs with a 2 vs. 6 comparison for which performance was above chance in Experiment 1. To test whether the lemurs were deciding based on odor cues, we pre-baited the bucket that was designated for the smaller quantity such that the two buckets provided the same olfactory cues.
4.3.1 Methods

4.3.1.1 Subjects

Subjects were 15 diurnal and catheval prosimians (8 females and 7 males; mean age 17.01 years, SD 9.53), housed at the Duke Lemur Center. Subjects represented a similar distribution of species as in Experiment 1. Thirteen out of 15 subjects were also tested in Experiment 1.

4.3.1.2 Procedure

The procedure was identical to Experiment 1 except that all trials consisted of a 2 vs. 6 comparison in which the bucket that was baited with 2 already contained 4 food items hidden in the bottom. This meant that when the baiting was complete, both buckets contained 6 food items providing the same olfactory cues\(^1\).

4.3.2 Results

Subjects chose the bucket into which they had observed six food items placed significantly more often than predicted by chance (12 out of 15 trials, \(p < 0.05\), one-tailed).

4.4 General Discussion

Our findings provide little support for the idea that prosimian primates use object-file representations to track food items. Instead the data are consistent with the

\(^1\) In contrast to Experiments 1 and 2, experimenters were not blind to the number of food items the other experimenter was baiting.
idea that lemurs spontaneously represent and compare quantities using the ANS. Subjects were able to successfully select the larger quantity with a 1:3 ratio but not a 1:2 ratio or a 2:3 ratio. The fact that lemurs were able to successfully discriminate 2 from 6 is also counter to the predictions of incommensurate representations. Failure to differentiate 2 from 3 food items further suggests that they were not benefitting from the increased precision offered by the object file system.

Hauser and others have suggested that the object-file system might be primary when animals engage in spontaneous numerical judgments without training and that extensive training might be required for animals to represent large values outside the purview of the object-file system (Hauser, et al., 2000; vanMarle, et al., 2006). However, we tested untrained animals in the same spontaneous cognition circumstances and found no evidence for the object-file system. These results emphasize the importance of selecting values that can directly contrast the predictions of two systems and test the limits of each system.

Not surprisingly, when animals are given a choice between two quantities of food where volume of food is equated, there is no motivation to select the larger numerosity and instead animals maximize total food intake (Stevens, et al., 2007; Beran, et al., 2008). In contrast, when duration has been controlled this appears to have no impact on choices and animals and babies have been shown to successfully choose the large (but shorter) sequence (e.g. Feigenson, et al., 2002; Hauser, et al., 2000). Given
these previous findings and the fact prosimians use many olfactory cues in other aspects of their lives, we chose to focus our limited resources on conducting a scent control rather than a volume or duration control. In Experiment 3, we gave lemurs a choice between buckets that were visibly baited with 2 or 6 food items, however we prebaited the bucket with 2 items with 4 additional items such that the scent cues were equalized. Despite this control, subjects selected the container that should have contained the larger number of food items based on the information with which they were provided.

The findings presented here should be considered in light of recent work directly comparing the numerical abilities of lemurs to monkeys, as well as comparing the numerical abilities of three different lemur species (Jones, et al., under review). In that study, rhesus macaques, ring-tailed lemurs, mongoose lemurs and blue-eyed black lemurs were trained to select the numerically larger of two visual arrays on a touch-screen. Results indicated an equivalent range of performance for all four nonhuman primate species. Despite the large variation in social structure, home range size, and diet in the species tested, Jones, et al. (under review) reported much more variability in the acuity of the ANS within each species than between species. We therefore find it unlikely that the lack of evidence for a set size limit in the spontaneous numerical comparisons of lemurs reflects a difference between prosimians primates and old world primates, but rather may reflect differences in factors that cue the two systems.
Recently, some authors have proposed that small quantities may be represented by both ANS and object file systems, and that contextual factors may determine which system is cued (Cordes & Brannon, 2009; Hyde, 2011). Recent work indicates that attentional load may influence the cueing of the two systems. A recent study showed that exact enumeration of small numbers (<4) is inhibited during a task with high attentional load, but approximate numerical representation is not (Burr, et al., 2010; but see Vetter, et al., 2008). Another study showed that individual differences in small number representation correlated with working memory, but ANS acuity does not (Piazza, et al., 2011). Hyde, et al. (2011) also suggested that spatial attention impacted which system will represent the numerical value of a small number set (1-3 items). Specifically, they report that when the spatial distribution of visual objects allowed for individuation, ERP responses showed a pattern consist with parallel individuation. In contrast, when attention could not select individual objects, ERP responses showed a pattern consist with ratio dependence. Hyde (2011) proposed a hypothesis that conditions that allow attentional selection of individuals cues the object file system, while conditions in which items are presented outside attentional limits, they will be represented as an approximate numerical representation. Our findings do not support this hypothesis: small quantities were presented sequentially without additional attentional requirements, and yet still resulted in approximate representations.
The only extant evidence for the object file system in nonhuman primates comes from spontaneous food choices like the task used here. However our results demonstrate that the general aspects of that task are insufficient to cue the object file system. Possible factors that could inhibit the ANS or cue the object file system must then relate to more specific aspects of the context in which such a task is performed. Candidates for these factors include, but are not limited to, satiation, inhibition, arousal, competition, dominance, and inadvertent social cues from the experimenter. Future research will need to address the contexts that cue the object file system in spontaneous discrimination tasks.

In sum, by employing conditions designed to specifically address the predictions of the ANS and the object-file system, we found that spontaneous numerical comparisons in prosimians primates are likely to be driven by the ANS. Given these results, the factors that may lead nonhuman primates to compare quantities using the object file system rather than the ANS remain unclear. What is clear, however, is that the ANS is spontaneously accessed by nonhuman primates to compare quantities regardless of the magnitude of those values being tested.
5. **Nonsymbolic, approximate multiplication in monkeys**

(Jones, S. M. & Brannon, E. M., in preparation)

A large literature demonstrates that nonhuman animals share with humans an ability to represent and compare numerical values approximately (see Brannon, 2005, 2006; Nieder & Miller, 2004 for reviews). The hallmark of the approximate number system (ANS) is ratio-dependence: accuracy in comparing two values depends on the ratio between them. Less research has addressed how representations from the ANS enter into arithmetic operations.

The ANS allows an appreciation of the ordinal relations between numerosities. Ordinality is perhaps the most basic form of mental arithmetic (Gelman & Gallistel, 1978). The ability to order numerical values requires the ability to appreciate the relationship between two values, rather than simply appreciating that they are nonequivalent. A variety of nonhuman species have been shown to select the larger of two values. Most studies of numerical discrimination in non-primates make use of the spontaneous preference to obtain the larger of two quantities of food items (e.g. Addessi, et al., 2008; Anderson, et al., 2007; Beran, et al., 2007; Hanus & Call, 2007; Uller & Lewis, 2009; vanMarle, et al., 2006) or to approach the larger of two set of conspecifics (e.g. Agrillo, et al., 2007, 2008). However, fewer studies have examined an abstract ability to numerically order sets of abstract stimuli which can be more carefully controlled for nonnumerical variables.
Brannon and Terrace (1998, 2000) trained rhesus macaques to order visual arrays of 1 to 4 items based on their numerosity. Monkeys were then tested on novel values 5 - 9. They found evidence that monkeys were able to transfer the ordinal rule, indicating that monkeys had not simply memorized pairwise associations between the small values, but had in fact learned a numerical rule. Cantlon and Brannon (2005, 2006b) extended these findings to include a bidirectional task in which rhesus macaques learned a set of ordinal rules, cued by the color of the display background. When the background was red, monkeys were rewarded for first selecting the numerically smaller of two arrays. When it was blue, they were rewarded for selecting the numerically larger array first. Monkeys were able to extend these numerical rules from a learning set to novel values (1 through 9). Cantlon and Brannon (2006b) further demonstrated transfer to larger values up to 30. There was qualitative and quantitative similarity in performance between the monkeys and a sample of adult humans: accuracy and reaction time were modulated by the ratio between values being ordered, and the numerical sensitivity for monkeys was within the range of sensitivity reported for humans.

The ability to numerically order sets of computerized stimuli has been demonstrated in a variety of other species, including capuchin monkeys Beran (2008a,c), chimpanzees (Tonomaga, 2008), lemurs (Jones, et al., under review; Merritt, et al., 2011),
and even pigeons (Scarff, et al., 2012), indicating that understanding ordinality as an abstract numerical rule might have evolutionarily ancient origins.

Humans and nonhuman primates have been shown to share a system of numerical representation that enables the ability to order, compare, and add magnitude representations. In addition to the basic arithmetic operation of ordering, humans show evidence for approximate addition (e.g. Barth, Beckmann, & Spelke, 2008; Cantlon & Brannon, 2007a; McCrink & Wynn, 2009; Pica, et al., 2004). Similarly, nonhuman primates have been shown to add quantities in an approximate way. In a food choice task, nonhuman primates reliably choose the larger of two quantities by tracking one-by-one additions to multiple caches (e.g. Beran, 2004; Beran, Evans, Leighty, Harris, & Rice, 2008; Hanus & Call, 2007; vanMarle, et al., 2006). These quantities are amassed sequentially such that they are never seen as a complete set. Thus, a subject’s representation of quantity in this task reflects the mental sum of all items.

In addition to one-by-one accumulation, nonhuman primates have also been found to calculate the sum of two sets. Flombaum, Junge, and Hauser (2005) demonstrated approximate addition in rhesus macaques using a violation of expectancy paradigm. Each monkey saw a set of lemons which was subsequently hidden by a screen. An additional set of lemons was then added to the occluded amount, resulting in a physically possible outcome of 4 or 8. Monkeys who saw an unexpected number of lemons when the screen was removed (4 when the expected sum should be 8, or vice
versa) looked longer than those who saw the expected number. Cantlon and Brannon (2007a) and Cantlon, Merritt and Brannon (under review) found similar approximate addition abilities in monkeys and adult humans. Cantlon and Brannon (2007a) showed rhesus macaques and adult humans two sets of dots presented sequentially on a computer screen. After this presentation, subjects were required to choose which of two simultaneous arrays was numerically equal to the sum of the sequentially presented sets. Cantlon and Brannon (2007a) reported that both humans’ and monkeys’ accuracy was ratio-dependent, suggesting that they added approximate numerical representations.

The ANS may support a precursor to mathematics in the form of approximate addition. Recently, it has also been suggested that the ANS may also support a precursor to multiplication (Barth, et al., 2009; McCrink & Spelke, 2010). Adult humans perform symbolic multiplication problems using rote memorization of verbal facts (e.g. Dehaene & Cohen, 1997; LeFevre, Bisanz, Daley, Buffone, Greenham, & Sadesky, 1996). However, multiplication can also be considered as a counting operation, in which the total quantity can be organized into groups of equal numerosity (Lampert, 1986). This model of multiplication is derived from the operation of addition. In this repeated addition model, multiplication is the result of the addition of multiple sets, each containing an equal numerical quantity. Thus, 3x4 can be calculated as the sum of 4 sets, each containing 3 items, or 3+3+3+3.
Multiplication as repeated addition differs from standard addition in that in a standard addition problem, one is given information about the value of each addend. In a multiplication problem, on the other hand, one receives only two pieces of information: the value of a single set, and the number of sets. Under the repeated addition model, the number of sets is taken as the operator; the numerosity of each set as the operand. An important assumption of this model is that the operator must be a whole number: it is much harder to intuitively conceive of adding .60 sets, each containing 5 elements, than it is to conceive of adding 5 sets, each containing 0.6 elements (0.6 + 0.6 + 0.6 + 0.6 + 0.6). When the operator is a whole number, multiplication necessarily results in a product that is larger than the operand.

Throughout early mathematical education, children rely on repeated addition to understand and calculate multiplication problems. For example, Lemaire and Siegler (1995) studied the acquisition of single digit multiplication by children in 2nd grade. Using a longitudinal design, they examined strategy use across a year in which children were learning single-digit multiplication. They reported that children used repeated addition to calculate the answers to the most difficult multiplication problems. This reliance on repeated addition continues even after mastery of memorized multiplication tables. Fischbein, Deri, Nello, & Marino (1985) reported that repeated addition impacted the understanding of multiplicative operations beyond childhood, up to adolescence, and thus argued that the implicit model for understanding multiplication is repeated
addition. They asked students in grades 5, 7, and 9 to choose the operation needed to solve word problems. Participants were worst at identifying multiplication as the appropriate operation when the assumptions of a repeated additions model was violated (i.e. when the operator was a decimal or fraction). This result indicates that despite a mastery of multiplication tables, adolescents were continuing to conceptualize multiplication as a process of repeated addition and were impaired in recognizing the operation of multiplication when it could not be readily conceptualized as such.

Repeated addition thus serves as a model of multiplication that may be accessible before mathematics training. Other types of multiplicative abilities have already been demonstrated in children and animals with no multiplication training. Barth, et al. (2009) and McCrink and Spelke (2010) both directly assessed approximate multiplicative abilities in young children who have not yet received mathematics training in multiplication. Specifically, they tested whether children were able to mentally transform approximate numerical representations. Participants were shown a static visual array of objects before it was occluded. During occlusion, they were told that the array was transformed by a “magic wand”, resulting in the numerical value of the array doubling, quadrupling, or increasing by a fractional factor (2.5). Participants were asked to compare their expectation of the outcome of this transformation to a separate, visible array. Participants were able to represent the outcome of the multiplicative transformation and showed the ratio dependent signature of the ANS.
Thus, there is evidence that approximate numerical representations can be multiplicatively transformed.

Models of animal conditioning and foraging often include the assumption that animals can estimate or calculate rates of reinforcement. Leon and Gallistel (1998), for example, investigated whether subjective magnitude of reward combined multiplicatively with rate of reward to determine a rat’s preference between two levers. They found an effect of the rates of reinforcement and of the magnitudes of the rewards. They also found that the factor by which preference increased in response change in rate was independent of magnitude. These results indicate a combinatorial process that the authors argued was a form of multiplication.

Measures of foraging time among various foraging sites also indicate that animals are able to represent rates of food occurrence (e.g. Baum & Rachlin, 1969; Herrnstein, 1961; Herrnstein & Loveland, 1975). A simple model can elegantly account for foraging behavior, a central postulate of which is that switching to a foraging patch is determined by the rate at which food has been observed to occur at that patch (Myerson & Miezin, 1980). Attempts to model this behavior without a calculation of rate have led to substantial difficulties (e.g. Lea & Dow, 1984). Thus, nonhuman animals appear to calculate rate, and may do so using a process similar to multiplication.

A few studies have investigated something akin to repeated addition in nonhuman primates. Beran, et al. (2005) provided preliminary evidence that nonhuman
primates may be able to mentally double numerical sets. Two chimpanzees and a rhesus macaque learned an association between container color and the number of food items in containers of that color, such that individual colors were symbolic representations of specific numbers of food items (i.e. all pink containers held 5 food items). Animals were offered a choice between a pair of similarly-colored containers on the one hand, and a set of visible food items on the other. All three subjects were able to maximize their rewards by mentally combining two nonvisible sets, which could either be an additive or a multiplicative process.

To a similar end, Addessi, et al. (2008) trained capuchin monkeys to associate tokens with different quantities of food. Like color, tokens acted as symbols by arbitrarily representing numerical value without having any iconic relation to their referent. Subjects were able to choose between combinations of tokens. Selecting the larger reward amount required that subjects perform either an additive or multiplicative calculation to represent the reward value of different token combinations. Capuchins were able to flexibly combine quantities to maximize their reward. In both of these studies, it is unclear whether subjects performed a form of mental multiplication on the value of a single set, or whether they were performing symbolic addition.

To investigate how approximate numerical representations may be used in a multiplicative calculation, we designed a task in which monkeys had to make inferences to calculate a repeated addition akin to multiplication. On each trial there were two
pieces of information are available: 1. the number of elements in each set contained, and 2. the number of sets. Monkeys were only shown the number of elements in a single set and had to infer the total number of elements based on the number of sets. Thus, here we examine whether animals are able to use the repeated addition model of multiplication to combine approximate numerical representations.

In investigating the multiplicative abilities of nonhuman primates, we are not seeking a precursor to the rote memorization that characterizes human adults’ performance on single-digit multiplication. Instead, we seek evidence that nonhuman primates are able to conceptualize and calculate repeated addition when it is framed as a multiplication problem. Specifically, we are interested in whether animals will make an inference that all sets contain the same numerosity, and use that information to calculate a sum. For the purposes of these experiments, we define simultaneous addition as a process of summing multiple visual arrays in which all sets of elements are simultaneously visible. We define multiplication as a process of summing multiple sets in the case that only the two pieces of information available are the numerosity of a single set and the number of sets.
5.1 Experiment 1

5.1.1 Methods

5.1.1.1 Subjects

Subjects were two female rhesus macaques, Feinstein and Boxer, with extensive prior numerical discrimination training, including nonsymbolic addition and subtraction (Cantlon, Merritt & Brannon, under revision; Cantlon & Brannon, 2007a). Subjects were tested in sound-attenuated booth while seated in a plexi-glass primate chair. Stimuli were presented on a touch-sensitive computer screen. Subjects initiated each trial by pressing a start stimulus, a red square presented in the bottom left corner of the screen. Following this response, a 2.45 second animated movie clip was presented which featured an addition or multiplication operation (Figure 12). After the clip ended, subjects were presented with two choice stimuli and were rewarded for selecting the stimulus that contained the numerical match to the sum of the two addends in the video clip. Correct responses elicited positive visual (light-up border) and auditory (chime) feedback and a (1.3ml to 2ml) Kool-Aid reward. Incorrect responses were followed by a warning tone and a 2 second timeout period. All trials were followed by a 1 second intertrial interval.
Figure 12: Example of 2+2 addition movie clip. Arrows indicate the direction of movement of the squares.

5.1.1.2 Stimuli

Movie clips consisted of between 1 and 4 green squares (each 332x274 pixels) positioned in the corners of the computer monitor. In addition clips, all green squares moved off the screen sequentially to reveal a white frame containing an array of stationary red dots underneath (Figure 12). In multiplication clips, only one square moved off screen and revealed dots (Figure 13). On all trials throughout the experiment the number of dots in each quadrant was identical such that the monkey could infer that
if she uncovered 2 dots beneath a green square and there were three squares the total
number of dots was six (2x3). Within each movie clip, dots were homogeneous and
consisted of one of three possible dot sizes: 625, 1225, and 1849 pixels. Consequently,
the cumulative surface area of visible dots in a single movie clip ranged from 625 to
14792 pixels. There were 18 unique movie clips used for each trial-type for each of the 5
training stages.

Figure 13: Example of 2x2 multiplication movie clip. Arrows indicate the
direction of movement of the squares.
The choice stimuli for all trials were trial-unique and were drawn by the Real-Basic program. The cumulative surface area of each choice stimulus was 2,000 or 10,000 pixels, such that both stimuli had the same area on half the trials and the correct product/sum had the smaller or larger area on 25% of trials respectively. A response strategy based on selecting the stimuli closest in cumulative surface area to the visible dots in the movie clip would result in chance levels of performance.

5.1.1.3 Procedure

For each trial the 1-4 occluders that appeared in the corners of the screen occluded the same number of dots. Training progressed in stages, starting with trials in which all occluders were removed to reveal all dots, and progressing to more complex trials that required inferential reasoning where only one occluder was removed to reveal the array. On these more complex trials subjects had to infer that the total number of dots present was the number of dots revealed under the occluder times the number of occluders. Subjects were tested 5 days per week, with 250 trials per session. They progressed through five stages of training before novel probe testing occurred.

5.1.1.4 Training

There were 5 different training stages and each stage was continued until performance exceeded 75% accuracy on all trial types for two consecutive days. For example stimuli and timing for different training stages, see Appendix B.
Addition Stage 1: 2+2 and 2+2+2+2. Movie clips began with 2 or 4 green squares. All squares moved off screen, each revealing an array of 2 dots. Choice stimuli consisted of an array of four dots and an array of eight dots and were trial-unique.

Addition Stage 2: 2+2 and 2+2+2+2. Movie clips were identical to those in Stage 1, with the exception that all squares moved back into their starting places after moving off screen for up to 1000 ms to reveal the dots. As in Stage 1, choice stimuli consisted of an array of four dots and an array of eight dots and were trial-unique.

Multiplication Stage 3: 2x2, 2x4, 1x4 and 4x2. Movie clips consisted of those from Stage 3 as well as novel trial types (1x4 and 4x2) where the first operand represents the number of dots per box and the second represents the number of boxes. On multiplication trials, only one green square moves off screen and returns to its starting place. As in Stages 1 - 3, choice stimuli consisted of an array of four dots and an array of eight dots and were trial-unique.

Multiplication Stage 4: 2x2, 2x4, 1x4 and 4x2 and numerical matching trials 1x1, 2x1 and 4x1. Movie clips consisted of those from Stage 4 as well as trials consisting of a single green square that moved to reveal 1, 2 or 4 elements. Choice stimuli consisted of all possible pairs of the correct answer paired with one of the other values {1,2,4,8}.

Mixed Multiplication and Addition Stage 5: 2x2, 2x4, 1x4 and 4x2 and addition trials 2+2, 2+2+2+2, 1+1+1+1 and 4+4. Movie clips consisted of those from Stage 4 as well as addition trials in which all squares moved off screen and returned to their starting
locations. These addition movie clips differed from those previously described only in that the upper left frame was not occluded at any point in the trial. There were 9 different exemplars of each of the trial types. Choice stimuli were identical to Stage 1 (i.e., 4 and 8).

5.1.1.5 Test Phase 1

The goal of the first test phase was to test the subjects’ ability to solve problems with familiar numbers of elements and a novel numbers of sets in nondifferentially reinforced trials. Subjects were given three sessions of Phase 1 test trials, which allowed a total of 185 probe trials. Familiar Multiplication and Addition trials were interspersed with nondifferentially reinforced novel probe trials in which familiar numerical arrays (1 or 2 dots) were presented in a novel number of green squares (3, for both addition and multiplication trials): 1x3, 2x3, 1+1+1 and 2+2+2. There were 9 different exemplars of each of the trial types. Choice stimuli for the critical probe trials consisted of an array of three dots and an array of six dots. Sessions consisted of 75% reinforced Mixed Multiplication and Addition trials, and 25% probe trials.

5.1.1.6 Test Phase 2

The goal of the second test phase was to test the subjects’ ability to solve problems with a novel numbers of elements and a familiar numbers of sets in nondifferentially reinforced trials. Subjects were given five sessions allowing a total of 323 nondifferentially reinforced probe trials. Mixed Multiplication and Addition trials
were interspersed with nondifferentially reinforced probe trials in which novel numerical arrays (3 dots) were presented in a familiar number of green squares (2 or 4): 3x2, 3x4, 3+3 and 3+3+3+3. There were 9 different exemplars of each of the trial types. Choice stimuli for the nondifferentially reinforced trials consisted of an array of six dots and an array of twelve dots. Sessions consisted of 75% reinforced Mixed Multiplication and Addition trials, and 25% probe trials.

5.1.2 Results

5.1.2.1 Trials to criteria

Boxer required 85 sessions to complete training Stage 1, and 14 sessions to complete training Stage 2. She has not progressed to testing, having completed 69 sessions of training Stage 3 without yet having reached criteria. All subsequent findings reported here will reflect only Feinstein’s performance.

Feinstein required 59 sessions to complete training Stage 1, 11 sessions to complete training Stage 2, 31 sessions to complete training Stage 3, 17 sessions to complete training Stage 4, and 37 sessions to complete training Stage 5\(^1\). After training, Feinstein completed an additional 19 Mixed Multiplication and Addition Stage 5 sessions interspersed with Test Phase 1 and Test Phase 2.

\(^1\) After Multiplication Stage 4, additional probe trials were run without any differential reinforcement. It became clear that Feinstein’s performance on the reinforced trials was poor during these sessions, so she continued training with Mixed Multiplication and Addition Stage 5. These earlier probe trials were excluded from analysis.
5.1.2.2 Feinstein accuracy during test phases

During Test Phases, Feinstein’s overall accuracy on standard trials (differentially reinforced familiar trials from the training phase) was above chance levels of performance (Test Phase 1: 89.91%; Test Phase 2: 87.70%). Her accuracy was above chance levels of performance regardless of whether the cumulative surface areas (SA) of the two choice arrays were equal (Test Phase 1: 89.79%; Test Phase 2: 87.29%), the SA of the reinforced array was larger than the SA of the distractor (Test Phase 1: 89.73%; Test Phase 2: 90.09%), or the SA of the distractor was larger than the SA of the reinforced array (Test Phase 1: 90.37%; Test Phase 2: 86.42%).

During Test Phase 1, Feinstein’s performance on the critical probe trials was significantly greater than chance expectation (Figure 14). For problems 1+1+1 and 1x3, she chose the answer 3 more often than predicted by chance (1+1+1: 97.67%, \( p < 0.001 \), two-tailed; 1x3: 89.83%, \( p < 0.001 \), two-tailed). For problems 2+2+2 and 2x3, she chose the answer 6 more often than predicted by chance (2+2+2: 72.09%, \( p < 0.01 \), two-tailed; 2x3: 87.50%, \( p < 0.001 \), two-tailed)). Accuracy on simultaneous addition probe trials did not differ from multiplication trials. Accuracy for 1+1+1 did not differ from accuracy for 1x3, \( \chi^2(1, N = 102) = 2.35, \text{NS} \). Likewise, accuracy for 2+2+2 did not differ from accuracy for 2x3, \( \chi^2(1, N = 83) = 3.08, \text{NS} \).
Figure 14: Feinstein was tested, without reward, on her ability to perform arithmetic calculations on probe trials in which familiar numerical arrays (1 or 2 dots) were presented in a novel number of green squares (3). Her performance indicated that she extended an arithmetic rule to novel stimuli for both addition and multiplication trials.

In Test Phase 2, Feinstein performed significantly above chance for two of the four probe trial types. She was above chance on problems 3x4 (95.29%, p < 0.001, two-tailed), and 3+3+3+3 (93.55%, p < 0.001, two-tailed). Her performance was significantly worse than chance on 3x2 (23.61%, p < 0.001, two-tailed), and was not different than chance for 3+3 (50.68%, NS). This pattern can be explained by a bias to choose 12 over 6 across the four probe trial types (80.19%, p < 0.001, two-tailed).

Feinstein’s preference for 12 over 6 across probe trials in Test Phase 2 does not reflect an overall bias to select the larger in standard trials. For problems 1+1+1+1, 1x4, 2+2, and 2x2, she chose the numerically smaller array more often than predicted by
chance (1+1+1+1: 91.80%, \( p < 0.001 \); 1x4: 86.55%, \( p < 0.001 \); 2+2: 88.03%, \( p < 0.001 \); 2x2: 67.37% \( p < 0.001 \)).

5.1.2.3 Testing alternative hypotheses

To assert that Feinstein was using both number of elements and number of sets to calculate an answer, and in effect performing a repeated addition variant of multiplication, I tested the following alternative hypotheses.

Alternative hypothesis 1: Feinstein was matching only the number of elements visible during each video. If this explained her pattern of responding, we would expect her to perform differently on the following reinforced trial types: 4+4 (in which all occluders are removed and all 8 dots are simultaneously visible) and 4x2 (in which only a single set is revealed). A pattern of responding based solely on the number of visible elements would result in Feinstein correctly picking 8 over 4 for 4+4, but incorrectly picking 4 over 8 for 4x2. Instead, she chose 8 over 4 more often than predicted by chance for both 4+4 trials (Test Phase 1: 98.75%, \( p < 0.001 \); Test Phase 2: 97.39%, \( p < 0.001 \)) and 4x2 trials (Test Phase 1: 93.55%, \( p < 0.001 \); Test Phase 2: 94.69%, \( p < 0.001 \)).

Alternative hypothesis 2: Feinstein learned an association between the number of elements and the reinforced response, while ignoring the number of sets (i.e. When there are 2 elements, choose 4 over 8.) If this explained her pattern of responding, we would expect her to perform identically on the following reinforced trial types: 2x2 and 2x4. Instead, Feinstein chose 4 over 8 more often than predicted by chance for 2x2 trials.
(Test Phase 1: 90.54%, \(p < 0.001\); Test Phase 2: 67.37%, \(p < 0.001\)), but chose 8 over 4 more often than predicted by chance for 2\(\times\)4 trials (Test Phase 1: 91.38%, \(p < 0.001\); Test Phase 2: 84.62%, \(p < 0.001\)).

Alternative hypothesis 3: Feinstein learned an association between the number of sets and the reinforced response, while ignoring the number of elements (i.e. When there are 4 sets, choose 8 over 4.) If this explained her pattern of responding, we would expect her to perform identically on the following reinforced trial types: 1\(\times\)4 and 2\(\times\)4. Instead, Feinstein chose 4 over 8 more often than predicted by chance for 1\(\times\)4 trials (Test Phase 1: 69.86%, \(p < 0.001\); Test Phase 2: 86.55%, \(p < 0.001\)), but chose 8 over 4 more often than predicted by chance for 2\(\times\)4 trials (Test Phase 1: 91.38%, \(p < 0.001\); Test Phase 2: 84.62%, \(p < 0.001\)).

Given these results, we can conclude that Feinstein learned to attend to both the number of elements per set and the number of sets for reinforced trials. To claim that she had learned an arithmetic rule, we must show successful transfer to novel probe trials. Her performance on the critical probe trials was significantly greater than chance expectation in Test Phase 1. To assert that Feinstein was showing transfer of an arithmetic operation, I tested the following alternative hypotheses.

Probe trial alternative hypothesis 1: Having memorized a series of “choose larger” or “choose smaller” response rules to standard trials, Feinstein mistook 3 sets for 4 sets. If this explained her pattern of responding on probe trials in Test Phase 1, we
would expect her to perform identically on the following trial types: 1x3 and 1x4, and 2x3 and 2x4. The proportion of 1x4 trials on which she selected the smaller array (chose 4 over 8: 69.86%) differed significantly from the proportion of 1x3 probe trials on which she selected the smaller array (chose 3 over 6: 89.83%), $\chi^2(1, N = 132) = 6.64$, $p = 0.01$, indicating that she did not confuse 3 occluders with 4 occluders. However, no such difference was found between the proportion of 2x4 trials on which she selected the larger array (chose 8 over 4: 87.50%) and the proportion of 2x3 probe trials on which she selected the larger array (chose 6 over 3: 91.38%), $\chi^2(1, N = 98) = 0.08$, NS.

Probe trial alternative hypothesis 2: Having memorized a series of “choose larger” or “choose smaller” response rules to standard trials, Feinstein mistook 3 elements for 4 elements. If this explained her pattern of responding on probe trials in Test Phase 2, we would expect her to perform identically on the following trial types: 3x2 and 4x2. The proportion of 4x2 trials on which she correctly selected the larger array (chose 8 over 4: 94.69%) differed significantly from the proportion of 3x2 probe trials on which she incorrectly selected the larger array (chose 12 over 6: 76.39%), $\chi^2(1, N = 185) = 11.90$, $p < 0.001$, indicating that she did not confuse 3 dots with 4 dots.

Probe trial alternative hypothesis 3: Having memorized a series of response rules resulting in a specific cardinal response to standard trials (i.e. when 2 elements and 4 sets, select an array of 8 elements), Feinstein mistook 3 for 4. This hypothesis is indistinguishable from transfer of an arithmetic rule in Test Phase 1. (If she confused
1x3 trials with 1x4 trials, she would select 3 as the choice array numerically closest to 4. Likewise, if she confused 2x3 trials with 2x4 trials, she would select 6 as the choice array numerically closest to 8.) However, this hypothesis is insufficient to explain Feinstein’s preference for 12 over 6 across all four probe trial types in Test Phase 2. (If she confused 3x2 trials with 4x2 trials, she would select 6 as the choice array numerically closest to 8.)

5.2 Discussion

The pattern of results reported here indicates that one monkey learned a numerical rule akin to multiplication that relies upon both the number of elements per set and the number of sets. To successfully perform on multiplication trials, Feinstein had to represent the number of elements in a single set and the number of occluders, and combine that numerical information multiplicatively. We argue that this operation reflects repeated addition, a model of multiplication in which a single numerical magnitude is added to itself multiple times. Importantly, she was also able to transfer this numerical rule to probe trials with novel numerosities. Alternative strategies were insufficient to explain her pattern of responses to standard trials or probe trials in Test Phase 1, providing strong evidence that the ANS supports a precursor to the arithmetic operation of multiplication.

Performance on probe trials in Test Phase 1 is not well explained by an alternative strategy of memorizing response rules specific to trial-types for standard trials and misapplying them to numerically similar probe trials. Specifically, the
proportion of correct responses for 1x3 trials differed significantly from the proportion of correct responses for and 1x4. Likewise, her accuracy on 4x2 trials differed significantly from her accuracy on 3x2 probe trials. This suggests that Feinstein’s behavior in response to probe trials differed from her behavior in response to standard trials. This data provides preliminary evidence that nonhuman primates are capable of the kind of repeated addition of mental magnitudes that underlies implicit models of multiplication in humans.

An overall preference for the numerosity 12 in Test Phase 2 complicates interpretation of these results. I posit that Feinstein’s performance on Test Phase 2 reflected a bias to select a value outside the range of choice stimuli with which she was trained. One possible direction for this line of research is to eradicate such a bias by training Feinstein on a broader range of values. It would be interesting to re-test the probe values from Test Phase 2 once the range of familiar values has been expanded to larger values. Although further testing is necessary to test the limits of the initial findings that Feinstein was able to transfer a multiplicative numerical rule, a range-based comparison strategy cannot explain her pattern of responding in Test Phase 1.

Although these results suggest an evolutionary precursor to the repeated addition model of multiplication, we must stress that these results are a far cry from the rote memorization practiced in single-digit multiplication by adult humans. Dehaene and Cohen (1997) provided evidence for a double dissociation between rote
memorization for multiplication and numerical calculation for other arithmetic operations. A patient with a selective deficit of verbal knowledge showed a pronounced deficit for multiplication, but not for addition or subtraction problems. In contrast, a patient who demonstrated a deficit in quantity processing was impaired in calculating responses to subtraction problems, but was able to recall memorized multiplication tables. This double dissociation between rote memorization and numerical calculation suggests that multiplication performance in adults is driven by verbally encoded facts.

On the other hand, repeated addition has been shown to be important to the conceptualization and calculation of multiplication problems. Grade school children rely on a strategy of repeated addition for difficult problems during acquisition of multiplication skills (Lemaire & Siegler, 1995). Humans continue to rely on a model of repeated addition to conceptualize and calculate multiplication through adolescence and early adulthood (Fischbein, et al., 1985; LeFevre, et al., 1996). The findings reported here suggest that the concept of multiplication as a process of repeated addition may originate with approximate numerical abilities shared with nonhuman primates.

An alternative model of multiplication has also been demonstrated using the approximate number system. McCrink and Spelke (2010) provided strong evidence that young children are able to perform scalar transforms on large approximate numerosities. Children were able to represent and transform a numerical value by doubling, quadrupling, or increasing that value by a fractional factor (2.5). This process
differs from repeated addition in that it is not constrained by the assumptions of the repeated addition model: the operator, which is repeated addition is operationalized as the number of sets, need not be a whole number. It is conceptually possible to scale an amount to 2.5x its starting value. It is not conceptually intuitive to add a quantity to itself 2.5 times. While we believe that the results reported here reflect the use of a repeated addition model of multiplication, it is possible that Feinstein was instead using the number of occluders as a perceptual cue to perform a scalar transformation. An example of such a scalar rule might be “When there are two occluders, scale the numerosity of the visible array by a factor of two.” In the case where the number of occluders is equivalent to the scalar value of the rewarded transformation, the results of these processes are identical. To discriminate between these two strategies for multiplication, future research could include a condition in which subjects are rewarded for using the number of occluders as a cue to perform a nonequivalent scalar transformation (i.e. “When there are four occluders, scale the numerosity of the visible array by a factor of 1.5.”) While this association is likely possible to train, I believe it to be less intuitive and predict that it would require significantly more training.

Another interesting topic for future research is the set limitation on repeated addition. Is there an upper limit on the number of times a quantity can be added to itself? In this task, we included 1-4 occluders. Recent research on the representation of numerical arrays by humans indicates a limit to the number of sets that can be
enumerated in parallel. Halberda, et al. (2008) presented human adults with visual arrays containing spatially overlapping dots of multiple colors. Participants were able to successfully enumerate subsets of dots on the basis of shared color, but showed a three-set limit in the number of color subsets that could be enumerated in parallel. This limit converges with three-element limits of parallel attention and visual short-term memory.

If the process of repeated addition requires that each numerically equivalent addend be represented simultaneously, a limit on parallel attention might also constrain the number of sets across which subjects can perform repeated addition using the ANS. For example, a subject might be able to simultaneously hold the numerosity of three arrays in working memory in order to mentally add them together (4+4+4 ≈ 12), but would be unable to hold the numerosity of six arrays in working memory and so would be unable to repeatedly add a value six times (4+4+4+4+4+4 = ?). Alternatively, if only the running total of a series of sequential additions is held in working memory, then no such limit need necessarily apply. In this case, a subject would not represent all sets in parallel, but would instead represent the approximate number of addition operations to perform using a single addend (4+4 ≈ 8 → 8+4 ≈ 12 → 12+4 ≈ 16 → 16+4 ≈ 20 → 20+4 ≈ 24).

Nonhuman primates have been shown to share with humans an approximate number system that can support the arithmetic operations of ordinality, addition and subtraction. The findings presented here support a growing literature indicating that the ANS is also able to support an approximate system of repeated addition that serves
as an implicit conceptual model of multiplication in adult humans. Evidence for the learning and transfer of this numerical rule to novel numerosities by a nonhuman primate suggests that, like other arithmetic abilities, multiplication may be an evolutionarily ancient cognitive ability.

6. Summary and Conclusions

Each of the studies presented here add important information to our understanding of how nonhuman primates represent number. Generally, these studies focused on the ability to discriminate between two visual arrays, while adapting the task to address context, precision, and species comparisons. Collectively, these studies show that the approximate number system likely evolved in primates as a context-dependent mental representation system that can be modulated by life experience. Such mental representations can be cued spontaneously and honed through training, and are common to both sequential and simultaneous presentation. This system may underlie the foundation of numerical representations and arithmetic operations in humans.

Chapter 2 provided evidence that acuity of the approximate number system is more characterized by individual differences than by genus or species differences in lemurs and macaques. It further indicates that an individual’s experience has an effect on the precision of their numerical representations. The results of Chapter 3 indicated that the cognitive process that monkeys use to compare numerical values is sensitive to the range of numerosities being presented, provided additional evidence for a shared
system of number representation in human and nonhuman primates. Chapter 4 tested the hypothesis that an object tracking system, rather than the ANS, might be invoked when animals engage in spontaneous numerical judgments without training. In contrast, the pattern of responding implicated the ANS in spontaneous quantity comparisons by nonhuman primates regardless of the magnitude of the values being tested. Chapter 5 expanded on the ability of the ANS to support arithmetic operations: results suggest that the ANS supports a process of repeated addition that serves as an evolutionary precursor to multiplication.

Ratio dependent performance in numerical discrimination is the hallmark signature of the approximate number system, and has been shown to be common to humans and nonhuman primates, as well as a variety of other species. Direct comparisons of the numerical abilities of humans with those of other species have provided strong evidence for qualitative and quantitative similarities between species (Beran, Johnson-Pynn, & Ready, 2008; Cantlon & Brannon, 2006b, 2007a; Jordan & Brannon, 2006a,b). The studies presented here add a richness and complexity to this picture: although the ANS is common to many species, the acuity of the ANS varies widely between individuals within a species. Further, the previous chapters suggest that the acuity of the ANS is malleable and dependent upon life experience, that the ANS is sensitive to range and context, and that nonhuman primates are able to compare, add, and even multiply mental magnitudes.
6.1 Chapter Summaries

Chapter 2 provided a comparison of multiple nonhuman primate species on a numerical comparison task with identical task parameters. This allowed for a direct comparison of two hypotheses: 1. That rhesus macaques would outperform lemurs on a measure of numerical acuity, as has been suggested by less direct comparisons, and 2. That ecological and social variables within lemur species may have led to different selective pressures on the evolution of numerical ability, which would manifest in species differences in ANS acuity. To this end, we tested rhesus macaques and three species of lemurs which differ on several ecological and social variables: blue-eyed black lemurs, ring-tailed lemurs, and mongoose lemurs. These three species differ with respect to diet, social structure and home range size (Table 3). These differences led to a series of predictions for selective pressures on acuity of the ANS: 1. If selective pressure to discriminate between quantities of discrete food items results in more precise number representations, mongoose lemurs and blue-eyed black lemurs would be expected to outperform ring-tailed lemurs by virtue of a due to an increased need to discriminate quantities of fruit. 2. If selective pressures to discriminate between the number of conspecifics result in greater numerical acuity, ring-tailed lemurs (and to a lesser extent, blue-eyed black lemurs) would be expected to outperform mongoose lemurs due to their more complex social structure. 3. If spatial abilities associated with
a large home range result in a more evolved magnitude representation system, the broad ranging ring-tailed lemurs should out-perform other species.

**Table 3: Ecological differences in three lemur species which might predict differences in numerical acuity at a species level**

<table>
<thead>
<tr>
<th>Ecological and social differences in three lemur species</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td><strong>Ring-tailed lemurs</strong> <em>(Lemur catta)</em></td>
</tr>
<tr>
<td><strong>Blue-eyed black lemurs</strong> <em>(Eulemur macaco flavifrons)</em></td>
</tr>
<tr>
<td><strong>Mongoose lemurs</strong> <em>(Eulemur mongoz)</em></td>
</tr>
</tbody>
</table>

Our sample sizes were too small to detect fine species differences. In fact, given the large individual variability, detecting species differences would require very large samples, beyond what is practical given the limited numbers of lemurs in captivity. Even species differences in numerical acuity between macaques and lemurs, if they exist, are not robust or gross enough to be detectable, due to individual variation. Hence, comparative research which attempts to characterize the numerical weber
fraction of a species using small sample sizes is likely to provide a poor representation of that species’ abilities. This concept of individual differences as important to comparative research is gaining popularity: in 2011, the International Journal of Comparative Psychology published a special issue entirely devoted to the role of individual differences in comparative psychology. Thus, one implication of Chapter 2 is a need for broader sample sizes and more cautious inferences.

Chapter 2 also tested the effect of experience on numerical acuity. Experiment 2 suggested that monkeys with extensive numerical training had more precise numerical representations than the monkeys who began Experiment 1A without any numerical experience. Experiment 3 supported the hypothesis that numerical experience affects acuity in that all three number experienced monkeys showed a reduction in weber fraction over number training. Future research should determine whether the gains in acuity from numerical experience transfer to different numerical tasks, or even result in the improvement of other nonnumerical cognitive tasks.

Chapter 3 showed a context effect on numerical comparisons using approximate representations, namely that monkeys, like humans, are sensitive to the range of values being tested. Subjects showed striking semantic congruity effects, replicating the findings of Cantlon and Brannon (2005). For at least one monkey, this semantic congruity effect was affected by the range of values, such that a given pair was considered small or large in relation to the numerical range. These results, combined
with evidence from Chapter 2 that both humans and nonhuman primates show ratio dependent performance, indicates that nonhuman primates and humans show quantitative and qualitative similarities in approximate numerical representations and the process of numerical comparison. Further, these results indicate that like adult humans, monkeys are capable of rapidly shifting their response strategies depending on context.
Chapter 4 addressed a different hypothesis regarding context-dependent response strategies. vanMarle and colleagues (2006) and Hauser, et al. (2000) both argued that extensive training may cause animals to represent quantity differently than they would do spontaneously, and suggested that in the context of a spontaneous food choice task, untrained animals may access the object file system rather than the ANS. However, I found no evidence for the object-file system in that context. Instead, the results provided strong evidence that nonhuman primates spontaneously access the ANS to choose between sets of food items. This finding indicates that a spontaneous food choice task is insufficient to cue the object file system in nonhuman primates.

These finding are in contrast to the results reported by Hauser, Carey and Hauser (2000), who found a set size limit in the spontaneous numerical discriminations of rhesus monkeys using a similar single-trial food choice task. As in Chapter 4, subjects watched as food items were dropped one-by-one into opaque containers, and were allowed to approach one box to retrieve food items. Since the single-trial food choice task itself is insufficient to elicit the object file system, our conflicting results indicate that contextual or methodological factors must cue the object file system over the ANS.

Chapter 5 showed that a nonhuman primate was capable of learning a numerical rule akin to multiplication that relied upon both the number of elements per set and the number of sets. The findings reported here suggest that, like humans, nonhuman primates are able to perform repeated addition of a mental magnitude as a model of
multiplication. This work expands on the findings that macaques are able to accumulate a mental magnitude through one-by-one addition of elements, and are able to mentally combine two sets to calculate a sum. Flombaum, Junge, and Hauser (2005) demonstrated the spontaneous addition of mentally represented sums in rhesus macaques. During a violation of expectancy viewing task, subjects saw a set of lemons which was subsequently hidden behind an occluder. An additional set of lemons was added to the occluded amount. Monkeys looked longer at the display if the screen dropped to reveal a set of food items that was numerically different than the sum of the sets they had witnessed. Cantlon and Brannon (2007a) and Cantlon, Merritt and Brannon (under review) also demonstrated approximate addition abilities in rhesus macaques. Cantlon and Brannon (2007a) showed subjects two sets of dots presented sequentially on a computer screen. Subjects were rewarded for selecting the sum of these sets from a pair of choice arrays.

Our current findings expand on this research by increasing the complexity of the task. Instead of simply combining two independently formed representations, subjects were required to infer that the same number of dots was present under multiple occluders. Unlike the addition tasks in which the representation of each addend was formed independently, subjects were required to represent the numerosity of a single array and to add that magnitude to itself repeatedly, using the number of occluders to determine how many times to repeat the operation. This results in an arithmetic
operation akin to multiplication. The successful learning of such a complex task requires that the subject use multiple types of numerical information and to combine representations multiplicatively.

Across the studies presented in the previous chapters, I have provided evidence for quantitative and qualitative similarities in the representation of number between lemurs and monkeys, and between nonhuman primate species and humans. The picture that emerges from these findings is the evolution of a cognitive ability common among primates, and likely many other species. Despite the ubiquity of the ability, the precision of the approximate number system appears to vary more widely between individuals within a species than it does between species. The ANS is recruited for spontaneous comparisons, but can be honed through experience, indicating both inherited and experience-dependent components. The result is a flexible cognitive system, able to represent numerosities for both sequential and simultaneous presentation across multiple formats, and yet sensitive to changes in context. This system may help to form the foundation for more complex numerical representations by supporting arithmetic operations.

6.2 Discussion

The findings reported in the body of this dissertation raise a number of interesting questions about the approximate number system. Here, I will discuss some of these issues and what new insight these studies provide. 1. How are numerical
discriminations affected by context? 2. What factors might cue the ANS or the object file system in a quantity discrimination task? 3. How can we characterize the evolution of the ANS?

6.2.1 Context

There is ample evidence that a variety of contextual factors relate to or impact approximate numerical discrimination. A sampling of these factors includes the spatial and temporal aspects of the stimuli or task, range of magnitudes, and task requirements. Here, I will review evidence that these factors interact with performance on tests of the approximate number system, and address how the findings reported in the previous chapters fit within this larger framework.

Many studies have demonstrated a relationship between numerical and spatial information, the SNARC effect being the most well-known (Dehaene, Bossini, & Giraux, 1993). Behavioral and neurological data indicate a relationship between representations of number and time (e.g. Gallistel & Gelman, 2000; Meck & Church, 1983), and between representations of number and space (e.g. Dehaene, Bossini, & Giraux, 1993; Hubbard, Piazza, Pinel, & Dehaene, 2005; DeWind & Brannon, 2012). Walsh (2003) proposed a theory of magnitude (ATOM) which hypothesized that time, space, and number are all represented by a common analog magnitude. In support of this hypothesis is evidence that the intraparietal sulcus (IPS), which is thought to play an important role in numerical processing (e.g. Dehaene, Piazza, Pinel, & Cohen, 2003), may not be domain-
specific for number (e.g. Cohen, Kadosh, Lammertyn, & Izard, 2008; Fias, Lammertyn, Caessens, & Orban, 2007; Ischebeck, et al., 2008). In line with their findings that the IPS is implicated in both numerical and alphabetical ordering, Fias and colleagues (2007) suggest that the IPS might be tuned to process ordinality.

Meck and Church (1983) provided preliminary behavioral evidence for a relationship between time and number. They found that rats spontaneously encoded both numerical and temporal information about a stimulus, and that their acuity for these discriminations was similar. If number, space and time are all represented using a common analog magnitude, we might expect the acuity of discrimination for these variables not just to be similar, as reported by Meck and Church (1983), but in fact to covary. DeWind & Brannon (2012) recently reported such a result for the numerical and spatial abilities of adult humans. They found that performance on discrimination of line length was positively correlated with performance on approximate numerical judgments. In a stronger test of the mutual interference of these parameters during discrimination, Merritt, Casasanto, & Brannon, (2010) recently found convincing behavioral evidence for a relationship between space and time processing. Rhesus macaques and human adults made ordinal judgments about the spatial dimension of stimuli while irrelevant temporal information varied, or vice versa. Results indicated that while both species showed an interaction between the two variables, only for monkeys was this interaction bidirectional. Thus, it is very likely that the temporal and
spatial properties of stimuli make up part of the context which influences numerical discriminations.

Other contextual factors that may influence numerical abilities include the magnitude of the stimuli, reference points, and the specific numerical tasks being utilized. Brannon, et al. (2006) found evidence that monkeys trained to order the numerical values 1, 2, 3, and 4 in descending order were unable to transfer the descending rule to pairs of larger novel numerical values. In contrast, monkeys trained to order 1-4 in ascending order successfully transferred the ascending rule to novel numerical values 5-9 (Brannon & Terrace, 1998, 2000). When trained to order values 4 – 6 in a descending order, however, subjects were able to extend the ordinal rule to smaller numerical values (1 – 3), but performed at chance levels with larger novel pairs (7 – 9). A monkey trained to respond in ascending order displayed the opposite pattern of results by accurately ordering novel values larger than the training set, but not smaller novel values. Brannon, et al. (2006) argued that monkeys establish the first value in a training sequence as a reference point and compare each subsequent numerical value with that reference value. Thus, the magnitudes of the stimuli and the order in which they are presented during training impact responding on a numerical rule-learning task.

Reference points may also be responsible for the semantic congruity effect in magnitude comparison tasks. Adult humans respond faster when the order of
responding (choose smaller vs. choose larger) is congruent with the relative magnitude of the stimuli. For example, Banks, et al. (1976) showed that adults chose the numerically larger of two relatively large digits more rapidly than they choose the smaller (e.g., 8 or 9) and vice versa. Cantlon and Brannon (2005) reported that monkeys show a similar semantic congruity effect when making approximate numerical comparisons. One model for semantic congruity effects highlights the role of reference points (e.g., Holyoak, 1978; Jamieson & Petrusic, 1975). These models assume that the values in a test pair are not directly compared. Rather, the relative distance between each value and a reference point determines response (Dehaene, 1989). The findings reported in Chapter 3 support this interpretation. Monkeys showed a semantic congruity effect which was modulated by the range of values being tested, as was previously reported in humans (Čech & Shoben, 1985). A given pair of values was treated as small or large depending on the range of values within the testing block in which it was presented. Further, subjects were capable of rapidly shifting their response strategies within the first few sessions after a contextual shift, presumably by shifting their reference point. Thus, the results of Chapter 3 indicate that the numerical semantic congruity effect in nonhuman primates is sensitive to contextual anchoring. Monkeys respond to identical stimuli differently depending on the context in which those stimuli are presented.
Finally, the findings presented here provided additional support that the specific numerical task can influence performance. In Chapter 3, the direction of the ordinal task and the magnitude of the stimuli combined to impact accuracy in a numerical discrimination task. Similarly, the results of Chapter 2 indicate that training history and task requirements can affect accuracy. Merritt, et al. (2011) trained ring-tailed lemurs to select the smaller of two arrays of dots (values 1 to 4). Lemurs transferred this ordinal rule to novel numerical values (5 - 9) replicating the basic pattern of results found with rhesus macaques (Brannon & Terrace, 1998, 2000). Merritt, et al. (2011) reported that lemurs’ accuracy was lower than that reported for macaques, although this finding was difficult to interpret because of differences in training. Specifically, lemurs were trained to order two sets of dots, while macaques had been trained to order the values 1 – 4. This difference in training may explain the accuracy difference reported between species. In Chapter 2, we trained both lemurs and macaques on a two-item descending ordinal task, and found overlapping response patterns. These results support the hypothesis that task requirements influence patterns of responding.

In Chapter 3, we refer to the context effect as the effect of numerical range on reaction time patterns during numerical comparisons. It is evident however that accuracy and reaction time on numerical discrimination tasks are influenced by a variety of contextual features. The temporal and spatial properties of stimuli, in addition to the range of values, can impact performance on a measurement intended to assess
numerical ability. Likewise, order of stimulus presentation and task requirements can modulate, or even eradicate, successful performance on a numerical task. The implications of this are clear: an individual’s ANS acuity is not invariant. An individual subject’s ability to discriminate a specific pair of numerosities depends upon the context in which that discrimination takes place.

6.2.2 Cuing the two systems

Context does not just affect numerical acuity, but may actually determine whether the ANS is recruited or inhibited for a numerical discrimination. Previously, researchers have argued that the object-file system might be accessed for spontaneous numerical judgments by untrained animals and that training or repeated testing might be required to recruit the ANS for large values (Hauser, et al., 2000; vanMarle, et al., 2006). Here, I examine the contexts in which these two systems have been documented for numerical discrimination tasks to examine which factors may be responsible for cuing one system over the other.

Strong evidence for the object file system requires showing successful discrimination of small values at a given ratio and simultaneous failure with large sets at the same ratio (e.g., 2 vs. 3 and 6 vs. 9). Several paradigms have provided evidence for the object file system in human infants. The food choice task used by Feigenson and colleagues, similar to the task described in Chapter 4, has repeatedly shown a set-size limitation in infants (Feigenson & Carey, 2005; Feigenson, et al., 2002). A manual search
task also demonstrated this set size limit in 10 to 14 month old infants (Feigenson & Carey, 2003, 2005). Set size limitations consistent with the object file system have been reported in a small set of nonhuman species. Nearly all of these studies rely on the food choice task described in Chapter 4. Hauser, Carey and Hauser (2000) demonstrated a set size limit in spontaneous numerical discrimination by rhesus macaques. Monkeys consistently chose the greater number of food items as long as the contents of each container did not exceed the set size limit of 4. Wood, Hauser, Glynn and Barner (2008) extended this finding to the discrimination of non-solid food portions. Beyond primates, set size limitations have been demonstrated using the food choice task with a small set of studies with other species. Uller and colleagues demonstrated a set size limit in discrimination of food quantities by horses (Uller & Lewis, 2009) and red-backed salamanders (Uller, et al., 2003).

A small number of studies with fish and birds have used a different paradigm to show the set size limit in quantity discriminations. When in a new environment, female mosquito fish will join a shoal, or collection, of other female fish for protection. Agrillo and Colleagues (2007, 2008) relied upon this natural tendency to examine the spontaneous quantity comparisons of fish and found the set size limit characteristic of the object file system, as well as ratio-dependent performance with large numbers. Rugani, et al. (2008) utilized a between-subjects training paradigm with newly-hatched
chickens. Chicks were successfully trained to discriminate 1 vs. 2 and 2 vs. 3, but failed to learn to discriminate larger values.

The majority of the experiments described above tested animals without extensive training using their spontaneous tendencies to select the larger quantity. The results reported by Rugani and colleagues (2009) stand out as the exception to this pattern. However, there is ample evidence that a lack of training is not by itself a sufficient context to reliably tap the object file system. Untrained animals have repeated shown ratio-dependent performance in spontaneous numerical discriminations (Flombaum, et al., 2005; Hauser, et al., 2003; Lewis, et al., 2005). From this we can infer that numerical training may eliminate a propensity to elicit the object-file system, but animals without number training also rely on the ANS in some contexts.

While these studies differ on a variety of factors (e.g. whether or not animals were separated from conspecifics for testing, whether stimuli were solid or nonsolid portions, whether objects were presented sequentially or simultaneously), the majority of studies that show a set-size limit in numerical comparisons share a few commonalities\(^1\). The quantities being discriminated were visual objects that moved independently from each other. This may align with the hypothesis that if individual elements do not move independently (e.g., are static or move as a unit), the object file

---

\(^1\) Here I exclude Rugani, et al. (2009) as an exception, both for using infant subjects, and for utilizing a between-subjects training design, both of which make their results difficult to interpret.
system may not be spontaneously activated (Barner, et al., 2008). Further, studies that show a set-size limit usually rely upon a natural preference for the larger amount, either of food or of conspecifics. Finally, a set size limit was exclusively found when at least one value in each numerical pair was smaller than the set size limit. In contrast, when Agrillo, et al. (2007, 2008) presented fish with a choice between two large values (8v12, 8v16), their discrimination showed the ratio dependence characteristic of the ANS.

By examining similar circumstances in which animals showed ratio-dependent performance, we can rule out some of these commonalities as being sufficient to cue to object file system. As we have seen, the demands of the task itself are insufficient to reliably cue the object file system. Lewis, Jaffe and Brannon (2005) found ratio-dependence in numerical comparisons by mongoose lemurs using a manual search task that was based on the paradigm used to show a set size limit in infants (Feigenson & Carey, 2003, 2005). The shoaling task used by Agrillo, et al. (2007, 2008) also elicits both the ANS and the object file system, although this may be a range effect: fish showed a set size limit when dealing with pairs of small quantities or a small and a large quantity, but showed ratio-dependent performance for larger values.

The food choice task itself is insufficient to cue the object file system, as are the use of only small numerical values. A number of studies using repeated presentations (as opposed to a single trial task) have previously shown ratio dependence in quantity discrimination in nonhuman primates (e.g. Addessi, et al., 2008; Anderson, et al., 2007;
Beran, Evans, Leighty, Harris, & Rice, 2008; Hanus & Call, 2007; vanMarle, et al., 2006). While many of these studies used pairs of values in which at least one exceeded the set size limit, vanMarle, et al. (2006) showed ratio dependence in the numerical comparisons of capuchin monkeys using small numbers only (1-4).

Evidence for ANS in these circumstances allow us to determine that several factors common among the findings of a set size limit are insufficient by themselves to cue to object file system or to inhibit the ANS. Evidence for the ANS has been found using untrained animals (e.g. Flombaum, et al., 2005; Lewis, et al., 2005), using visually presented objects that moved independently from each other (e.g. Agrillo, et al., 2007, 2008; Lewis, et al., 2005; vanMarle, et al., 2006), using tasks that rely upon a natural preference for the larger amount (e.g. Addessi, et al., 2008; Anderson, et al., 2007; Beran, Evans, Leighty, Harris, & Rice, 2008; Hanus & Call, 2007; vanMarle, et al., 2006), and using only small values (e.g. vanMarle, et al., 2006). Thus, none of these factors on their own are sufficient to recruit the object file system over the ANS. In Chapter 4, I was able to test whether these three factors combined in a single-trial food choice task would be sufficient to recruit the object file system over the ANS in untrained nonhuman primates.

The food choice task described in Chapter 4 provides a spontaneous measure of discrimination in which food items were sequentially presented. Regardless, the results were characterized by ratio dependence and not a set size limit. Here, I consider other
factors that differed between this task and the task from Hauser, et al., (2000). These include the range of values chosen, the testing conditions of the subjects, and controls for experimenter bias. Each of these differences provides factors that might influence what kind of mental representation is accessed.

Range of values - As shown in Chapter 3, nonhuman primates are sensitive to the range of numerosities being presented and may alter their behavior depending on this context. However, I used a single-trial task, so no subject was exposed to the entire range of values. Thus, the subjects tested in conditions which paired two small values can be considered separately as discriminating only small values. If the use of small values only in this task were sufficient to cue the object file system, we would predict the successful discrimination of 2 vs. 3 food items. As this was not the case, we can conclude that while large values may cue the ANS, small values do not necessarily cue the object file system.

Experimenter bias – In the majority of quantity discrimination research in which experimenters physically present the stimuli, the experimenters have been aware of which container held the larger quantity, and in some cases, of the experimental hypotheses. This allows for the possibility that animals may perform based on unintentional social cues from the experimenters. Our simple modification allowed the experimenters presenting food items to be blind to the condition on any given trial. Future research would benefit from including such designs, as there is ample evidence
outside of Clever Hans that nonhuman animals are able to read human social cues (e.g. McKinley & Sambrook, 2000; Miklo’si & Soproni, 2006; Tomasello, et al., 2001). However, given the findings of Agrillo, et al. (2007, 2008) in which stimuli were groups of fish in adjacent tanks and were thus experimenters did not physically present the stimuli, we can conclude that both systems can be cued without experimenter bias.

Testing conditions – In the studies presented in Chapter 4, subjects were separated briefly from conspecifics during testing to reduce interruptions and competition. This stands in stark contrast to the subjects from Hauser, et al. (2000) and Wood, et al. (2008), who were free-ranging animals that could not be isolated in separate enclosures. Wood, et al. (2008) reported that trials had to be aborted if the subject was distracted by another individual during the presentation or choice period. This highlights a critical difference between test settings: in Hauser, et al. (2000) and Wood, et al. (2008), animals were subject to interference by other animals, while no such interference was an issue at the Duke Lemur Center. Two potentially important factors lead logically from this contextual difference, namely differences in attentional load and social competition. Recent work has indicated that attentional load may influence the cueing of the two systems (Burr, et al., 2010; Hyde, et al., 2011). Hyde (2011) proposed that conditions which allow attentional selection cue the object file system, while conditions in with higher attentional load or inhibited spatial attention result in a reliance on approximate numerical representation. However, our findings may
contradict the predictions of this hypothesis: the higher attentional load required to track items in Hauser, et al. (2000) and Wood, et al. (2008) led to the object file system, while the lower attentional requirements of animals tested in isolation in our task led to approximate numerical representations.

In conclusion, there are likely a number of contributing factors that combine to recruit the ANS or the object file system in different contexts. What is clear is that no single contextual factor has been shown to be sufficient to recruit the object file system. The fact that only one study involving numerically trained animals has resulted in a pattern of responding consisted with the object file system (Rugani, et al., 2009) indicates that a spontaneous measure and a lack of numerical training may be necessary but not sufficient for an adult animal to demonstrate a set size limit in quantity comparisons. Other factors that may influence results are range of values tested, stimulus properties, and attentional requirements. However, while some or all of these contextual factors might be necessary for to recruit the object file system or inhibit the ANS in a numerical discrimination task, none of them in isolation are sufficient to do so.

### 6.2.3 Evolution of the ANS

Primates share quantitative and qualitative similarities in the representation of number. Many species, including prosimian primates, share a common mechanism for representing number approximately, and the acuity of that system overlaps between species. It is reasonable, then, to consider this system to be evolutionarily ancient. The
fact that acuity varies so widely between individuals within a species and is malleable through training indicates that there is also a learned component to the approximate number system.

Selective pressures for the existence of the ANS seem to have been so extreme and ubiquitous that there is little variation across species. Indeed, several recent studies have provided evidence for similar numerical abilities in incredibly diverse species. Scarf, et al. (2011) demonstrated that monkeys and pigeons are both able to acquire an abstract numerical rule and apply it to novel numerosities. Agrillo, et al. (2012) also reported qualitative similarities in numerical representation systems between phylogenetically distant species. They compared the numerical discrimination ability of humans and guppies using small and large value and found evidence that two distinct systems underlie quantity discrimination in both humans and fish. There are two potential explanations for this: 1. numerical abilities may have evolved independently in different taxa through convergent evolution, or 2. common mechanisms for numerical representation may exist between primates, birds and fish, suggesting that the evolutionary emergence of numerical abilities may be very ancient and common to a vast number of species. This poses an interesting dilemma. If numerical abilities have emerged through convergent evolution, for which species would we predict a lack of selective pressure to discriminate between quantities?
Despite the ubiquity of approximate numerical abilities, there is considerable evidence that training and individual experience shape the precision of the ANS. If the ANS is a cognitive mechanism evolved under strong selective pressure and closely related to fitness, individual variability and malleability may be surprising. However, a large number of genetic and environmental events affect complex, behavioral traits. This may result in higher genetic and nongenetic variability in traits most closely related to fitness (e.g. Houle 1991, 1992). The individual variability and improvement across training reported here may suggest that, in addition to heritability, environmental factors and experience within the lifespan play an important role for the approximate number system.

6.3 Conclusion

The studies presented here add rich information to our understanding of approximate numerical abilities in nonhuman primates. The evidence reviewed here indicates that the approximate number system evolved in primates as a context-sensitive, abstract mental representation system that can be modulated by experience within the lifespan. The ANS is recruited for spontaneous comparisons and trained discriminations, and can be honed through training, indicating both inherited and experience-dependent components. Approximate numerical representations underlie a variety of arithmetic abilities, from a simultaneous ordinal discrimination to the manipulation and multiplication of a remembered magnitude. It is unclear whether this
flexible cognitive system, which is able to represent quantity across multiple contexts, evolved separately in multiple species or is so evolutionarily ancient as to predate the divergence of the two basic forms of vertebrates, fishes and tetrapods. Future research is needed to test the limits of the ANS, both comparatively and contextually, and to explore the relationship between the arithmetic operations supported by the ANS and more complex symbolic mathematics.
Appendix A

Table 4: Chapter 4, Experiment 1 included a wide range of prosimian species, equally distributed between the numerical conditions.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Genus</th>
<th>Species</th>
<th>N</th>
<th>Sex</th>
<th>Mean age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eulemur</td>
<td>albibrons</td>
<td>4</td>
<td>2 females</td>
<td>25.64 years (SD 3.63)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>collaris</td>
<td>5</td>
<td>3 females</td>
<td>23.69 years (SD 5.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>coronatus</td>
<td>6</td>
<td>4 females</td>
<td>15.65 years (SD 8.62)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fulvus</td>
<td>1</td>
<td>1 female</td>
<td>26.64 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>macaco flavifrons</td>
<td>8</td>
<td>3 females</td>
<td>10.84 years (SD 9.91)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>macaco macaco</td>
<td>3</td>
<td>1 females</td>
<td>25.17 years (SD 3.02)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mongoz</td>
<td>8</td>
<td>4 females</td>
<td>15.31 years (SD 8.31)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rubiventer</td>
<td>7</td>
<td>3 females</td>
<td>22.35 years (SD 4.04)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rufifrons</td>
<td>1</td>
<td>1 male</td>
<td>29.03 years</td>
</tr>
<tr>
<td></td>
<td>Hapalemur</td>
<td>griseus</td>
<td>5</td>
<td>4 females</td>
<td>14.24 years (SD 1.93)</td>
</tr>
<tr>
<td></td>
<td>Lemur</td>
<td>catta</td>
<td>31</td>
<td>22 females</td>
<td>7.60 years (SD 7.89)</td>
</tr>
<tr>
<td></td>
<td>Propithecus</td>
<td>coquereli</td>
<td>16</td>
<td>8 females</td>
<td>8.65 years (SD 7.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>diadema</td>
<td>1</td>
<td>1 male</td>
<td>18.26 years</td>
</tr>
<tr>
<td></td>
<td>Varecia</td>
<td>rubra</td>
<td>11</td>
<td>4 females</td>
<td>11.18 years (SD 9.77)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>variegate</td>
<td>6</td>
<td>2 females</td>
<td>11.98 years (SD 12.09)</td>
</tr>
<tr>
<td>2</td>
<td>Eulemur</td>
<td>collaris</td>
<td>1</td>
<td>1 male</td>
<td>18.63 years</td>
</tr>
<tr>
<td>Species</td>
<td>Population</td>
<td>Sex Distribution</td>
<td>Average Age (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>-----------------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>coronatus</em></td>
<td>1</td>
<td>1 female</td>
<td>16.87 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>macaco flavifrons</em></td>
<td>3</td>
<td>2 females, 1 male</td>
<td>7.95 years (SD 11.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>macaco macaco</em></td>
<td>1</td>
<td>1 male</td>
<td>23.09 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>mongoz</em></td>
<td>2</td>
<td>2 females</td>
<td>13.37 years (SD 17.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hapalemur</em></td>
<td>1</td>
<td>1 female</td>
<td>13.50 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lemur</em></td>
<td>5</td>
<td>4 females, 1 male</td>
<td>5.57 years (SD 3.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Propithecus</em></td>
<td>3</td>
<td>1 female, 2 males</td>
<td>11.54 years (SD 5.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Varecia</em></td>
<td>2</td>
<td>1 female, 1 male</td>
<td>15.51 years (SD 16.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>variegata</em></td>
<td>1</td>
<td>1 female</td>
<td>7.00 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Eulemur</em></td>
<td>1</td>
<td>1 female</td>
<td>31.09 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>collaris</em></td>
<td>2</td>
<td>2 females</td>
<td>19.19 years (SD 4.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>coronatus</em></td>
<td>1</td>
<td>1 female</td>
<td>24.10 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>macaco flavifrons</em></td>
<td>1</td>
<td>1 male</td>
<td>1.22 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>macaco macaco</em></td>
<td>2</td>
<td>1 female, 1 male</td>
<td>24.71 years (SD 3.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>mongoz</em></td>
<td>3</td>
<td>2 females, 1 male</td>
<td>21.30 years (SD 5.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hapalemur</em></td>
<td>1</td>
<td>1 female</td>
<td>17.24 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lemur</em></td>
<td>2</td>
<td>1 female, 1 male</td>
<td>5.10 years (SD 2.82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Propithecus</em></td>
<td>2</td>
<td>2 males</td>
<td>9.82 years (SD 10.41)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Figure 15: Chapter 5, Experiment 1. Addition Stage 1: 2+2
Figure 16: Chapter 5, Experiment 1. Addition Stage 1: 2+2+2+2
Figure 17: Chapter 5, Experiment 1. Addition Stage 2: 2+2
Figure 18: Chapter 5, Experiment 1. Addition Stage 2: 2+2+2+2
Figure 19: Chapter 5, Experiment 1. Multiplication Stages 3 through 5: 2x2
Figure 20: Chapter 5, Experiment 1. Multiplication Stages 3 through 5: 2x4
Figure 21: Chapter 5, Experiment 1. Addition Stage 5: 2+2
Figure 22: Chapter 5, Experiment 1. Addition Stage 5: 2+2+2+2
References


resolving the evolutionary history of Madagascar’s lemurs. *Genome Research, 18*(3), 489-499.


McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition, 116*(2), 204-216.


Biography

Sarah Mychal Jones was born in Northampton, MA, on August 26, 1981. She attended Smith College (1999-2003), earning a Bachelor of Arts degree, with a double major in Psychology and Mathematics. Sarah was awarded a postbaccalaureate intermural research training award (IRTA) with the National Institutes of Health (NIH) in 2005. She was an IRTA fellow in the Laboratory of Comparative Ethology at the National Institute of Child Health and Human Development from 2005 to 2007. Sarah began graduate school in the Department of Psychology and Neuroscience at Duke University in the fall of 2007 where she was awarded the James B. Duke Award.

Publications


