# Numerical Cognition in Adults:
# Representation and Manipulation of Nonsymbolic Quantities

by
Hilary C. Barth

B.A., Psychology
Bryn Mawr College, 1996

Submitted to the Department of Brain and Cognitive Sciences in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy in Cognitive Neuroscience

at the

Massachusetts Institute of Technology

February 2002





Signature of Author: ________________________________________

Department of Brain and Cognitive Sciences
October 1, 2001

Certified by: ________________________________________

Nancy Kanwisher
Professor of Cognitive Neuroscience
Thesis Supervisor

Certified by: ________________________________________

Elizabeth Spelke
Professor of Psychology
Thesis Supervisor

Accepted by: ________________________________________

Earl Miller
Class of 1956 Associate Professor
Chairman, Department Graduate Committee

## Acknowledgments

Thanks to:

My parents, Karen and Frederic Barth, for everything.

Richard Gonzalez, who is really the reason I ended up majoring in Experimental Psych in the first place; Earl Thomas, who got me interested in neuroscience as well; and J. Toby Mordkoff, my undergraduate advisor after I decided to work with humans instead of pigeons or rats, who passed on some excellent training in cognitive science and good-natured cynicism.

Ruth Jennison, Elizabeth Blake, Alexis Perkins, and my other friends from Bryn Mawr for years of friendship and heated discussion. Thanks also to my pretty and popular friends from Boston as well (too many to list, but you know who you are), for being one of the most impressive groups of people I have ever had the good fortune to encounter. Special thanks to the rest of Pihaus.

Sean Gomez, who is part of the above group but deserves special thanks as well, for being there through the final steps in this process; and James Graham, for the same throughout the earlier years of grad school.

Many BCS people, including past and present members of Spelke lab, Kanwisher lab, and the Perceptual Science Group, for comments, discussion, and everything else, and Denise Heintze, for straightening out my administrative tangles over the years.

My committee members Randy Gallistel, for joining in at the end of a rather hectic process, and Mary C. Potter, for advice throughout the last five years.

And especially thanks to my advisors. Nancy Kanwisher, for agreeing to help me jump into a completely different area, halfway through the program, and for her excellent advice and support. Liz Spelke, for the class that made me want to think about this topic in the first place, and for her encouragement throughout the project. I count myself lucky to have had the chance to work with two such brilliant scientists.

Table of Contents

# Chapter 1

# General Introduction

How we acquire and represent knowledge is one of the most basic concerns of cognitive scientists. Certain lines of inquiry within the field have proven extremely useful to the exploration of these issues. The study of numerical cognition is undoubtedly one of these areas, as it has provided a richer understanding of domain specificity, biological determination of knowledge domains, the nature of conceptual change, the kinds of knowledge that may be present in beings without language, and the building blocks of that most remarkable of human achievements, higher mathematics.

Until fairly recently, researchers and laypersons alike probably would have been reluctant to attribute numerical abilities to infants or animals. The Clever Hans incident at the turn of the century, in which prominent scientists of the day were convinced that a horse had learned advanced mathematics, bred a strong resistance to the concept of numerical competence in animals. The issue remains controversial now as well, but evidence continues to build that preverbal human infants and nonverbal animals possess considerable, and sometimes quite surprising, numerical abilities. Results from studies on nonsymbolic number processing in human adults (for example, estimating the approximate number of elements in a large group) have shown remarkable parallels to the comparative work on other species. Indeed, infant, animal, and human adult numerical abilities share such similar characteristics that many researchers have supposed that a domain-specific system of knowledge, present in many species, is responsible for the sense of number and forms the basis for the complex symbolic manipulation of number developed by humans alone (Gallistel and Gelman 1992; Dehaene 1997). Though recent progress in this area has been considerable, many unanswered questions remain about the nature of number representation in humans and animals. This thesis addresses three of these questions. To what extent is an approximate mental magnitude independent of the form of the enumerated stimulus? Are these magnitudes constructed from different sorts of stimuli through a unitary process, or does the specific enumeration process depend on the sensory qualities of the stimulus set? And if more sophisticated symbolic mathematics is built upon the number sense, can this sense in its basic, nonverbal form be used in simple mathematical operations? The following sections will consider what is known about the number sense, the representations that underlie it, and its use in arithmetic computations.

**1.1 What does it mean to have an abstract "sense of number?"**

An animal (or a human) can be said to possess a concept of number only if certain criteria are met. The most basic requirement is that the animal must be shown to be able to base its behavior on numerosity unconfounded from the continuous variables that tend to covary with numerosity. For example, one might demonstrate that a rat can discriminate 4 dots from 8 dots, but this would not necessarily mean that the rat represented the number of dots. If the dots were all the same size, the rat might have made its discrimination based on the overall spatial extent of the stimuli. If the stimuli were sequentially presented, temporal extent, rather than spatial, would tend to covary with numerosity. Judgments based on these continuous variables provide evidence of representation of *quantity*, a continuous measure, but not of *number*, a discrete measure.

To demonstrate that numerosity can modulate behavior, however, does not provide sufficient evidence of a true number concept. Number is a highly abstract property of a set. It is independent of the identity of the set's elements, the arrangement of the elements, or any other features of the set. Once they have learned some counting techniques, human adults and children, can recognize the similarity between a group of fourteen apples and a series of fourteen beeps, though the sets may share no characteristics besides numerosity. This is because our system of symbolic enumeration allows us to label each group with a numeral that captures this abstract property of numerosity. Symbolic numerals may apply to any type of set; they are perfectly abstract. Therefore if we as human adults want to represent "how many" apples we have, for example, we do not need to retain a visual representation of a group of objects – we can simply remember "14." Similarly, if we want to represent a very different kind of set, such as a temporally extended series of beeps, we do not have to retain a memory of the entire sequence – we can again remember "14." Imagine, though, how we might represent "how many" *without* the help of a symbolic number system. One possible way to do this would be to form a visuospatial representation when dealing with apples, and a different sort of representation based on time, rather than space, when dealing with beeps. This particular method would *not* allow its user to recognize the numerical similarities between the sets of apples and beeps; it would *not* constitute an abstract number sense. Clearly, it is possible to represent a quantity in this manner; therefore, to demonstrate that animals or infants can respond based on a set's numerosity does not necessarily imply that they possess an abstract sense of number. The critical further requirement is to show that their responses generalize across numerosity when other properties change. This requirement also holds for human adults performing approximate number tasks that preclude the use of Arabic symbols. For the representation of magnitude to be considered truly abstract, it must not be limited to a particular sensory modality or presentation format.

Without a symbolic system to free its user from the constraints of set identity or sensory modality, then, could abstract number be represented? A large body of experimental work suggests that nonlinguistic creatures do in fact represent number, and some of this work strongly suggests that this sensitivity is truly numerical and abstract.

*Numerical capabilities in nonhuman animals*

Many kinds of nonhuman animals, ranging from pigeons and rats to monkeys and chimpanzees, have been found to make discriminations based on the numerosities of sets of various sizes. This is true for sets presented simultaneously or sequentially, visually or auditorily (Davis and Perusse 1988; Gallistel 1990; Gallistel and Gelman 1992; Dehaene 1997; Dehaene, Dehaene-Lambertz, & Cohen 1998). In one elegant study, rats were first trained to discriminate between an 8-second sequence of 8 tones and a 2-second sequence of 2 tones. After the discrimination was learned, the rats were able to respond with the "8" lever for the former sequence and the "2" lever for the latter sequence. Then test trials were presented in which either duration or number varied. For example, when duration varied, each sequence contained 4 tones, but its duration could be anywhere from 2 to 8 seconds. When number varied, on the other hand, each sequence was 4 seconds long, but it could contain anywhere from 2 to 8 tones. These test sequences allowed the researchers to assess whether the original discrimination had been based on duration or number. The results showed that the rats had kept track of both duration *and* number; when duration was fixed, the rats responded based on number, but when number was fixed, they responded based on duration (Meck and Church 1983). This study provides conclusive evidence that number, not a covarying continuous quantity, dictated the rats' behavior.

Animals have been found to generalize across stimuli on the basis of numerosity, independent of stimulus shape, color, identity, or modality; these results provide strong evidence for abstract numerical ability (Gallistel and Gelman 1992; Dehaene, Dehaene-Lambertz, & Cohen 1998). One of the most striking examples of the use of modality-independent numerical information comes from rats (Church and Meck 1984). The animals were trained to press one lever when presented with two lights or white noise bursts, and another lever when presented with four lights or white noise bursts. When two lights and two sounds were presented together, the rats pressed the "four" lever, suggesting that they spontaneously combined the quantities of light and sound and responded to their sum. This occurred even though the compound stimulus was a combination of two other stimuli, each of which taken alone demanded a different response. When a previously unseen stimulus was used (one sound and one light) the rats again responded to the sum, pressing the "two" lever.

*Numerical capabilities in human infants*

In a typical habituation experiment, infants are exposed to particular numerosities until their attention flags. Then they are presented with either that same numerosity or a novel one, and if they dishabituate to the novel numerosity, but not the original, it is taken as evidence of numerical discrimination. Such experiments have shown that infants of various ages can discriminate between 1 and 2 and between 2 and 3, for stimuli that are objects of various sizes, shapes and identities (Starkey and Cooper 1980; Antell and Keating 1983; Starkey, Spelke and Gelman 1983), and also for pronounced syllables (Bijeljac-Babic, Bertoncini and Mehler 1991), or for sequentially presented events (Wynn 1996). Infants can also discriminate 8 from 16 sounds, or 8 from 16 dots, though they fail at 8 vs. 12 (Lipton and Spelke 2000; Xu and Spelke 2000). A different technique presents infants with pantomimed additions or subtractions of very small numbers of objects. These simple arithmetic studies have demonstrated that infants are aware of the number of objects that should be present after the addition or subtraction: they show surprise when the resulting number is the incorrect one (Wynn 1992; Simon, Hespos, and Rochat 1995; Koechlin, Dehaene, and Mehler 1998; Uller, Carey, Huntley-Fenner, and Klatt 1999). Some, though not all, of these studies incorporated controls for non-numerical stimulus properties, ensuring that infants can respond to number and not some other covarying quantity.

Can infants relate the numerosities of sets presented to different sensory modalities? Looking times in six- to eight-month-old infants have been found to depend on the correspondence between the number of objects in a visual array and the number of drumbeats in a sequence (Starkey, Spelke and Gelman 1983; Moore, Benenson, Reznick, and Peterson 1987; Starkey, Spelke and Gelman 1990). However, it is possible that the infants simply matched objects to drumbeats through 1:1 correspondence, which does not require recognition of numerosity. Also, a recent replication attempt found no preference in either direction (Mix, Levine and Huttenlocher 1997). In addition, 3-year-old children do not perform well on crossmodal numerosity matching tests, in which they must choose the visual-spatial array that corresponds in number to a sequence of sounds (Mix, Huttenlocher and Levine 1996). While it is certainly possible that infants possess modality-independent numerical ability, to date evidence of this ability is equivocal.

*Numerical capabilities in human adults*

Though the case for abstract, purely numerical representations in human infants and nonhuman animals is still not entirely conclusive, there is a great deal of convincing evidence for the existence of a true number concept in adults. The characteristic patterns of performance on nonsymbolic number tasks with human adults closely follow the results of similar tests on animals, a parallel which has led many investigators to believe that its source is a common biologically determined system. One important shared feature of animal and human data is the fact that numerical discrimination is subject to Weber's Law, which states that the discriminability of two magnitudes depends on their ratio. For example, when rats are trained to press a lever a certain number of times before they can receive a reward, the number of presses produced is more accurate when the standard number is smaller. The rats become less accurate when they are comparing the produced number of presses to a larger standard (Mechner 1958; Mechner and Guevrekian 1962; Platt and Johnson 1971). The same is true for human adults whether they are producing lever presses, comparing numbers of dots, or even comparing Arabic digits (Moyer and Landauer 1967; Dehaene, Dupoux and Mehler 1990; Allik and Tuulmets 1991; Dehaene 1997; Whalen, Gallistel, and Gelman 1999). This last finding, that *symbolic* numerical comparisons are subject to Weber's Law, suggests that preverbal numerical representations come into play automatically, even when the task deals only with Arabic numerals. This finding is encouraging to researchers who wish to study the *nonlinguistic* aspects of quantity processing that humans share with other species. One might imagine that any human adult performing a numerical task might make use of his/her learned verbal knowledge of numbers, which could produce difficulties for researchers attempting to isolate preverbal numerical ability. This concerned is addressed to some extent by the fact that there is convincing evidence for such approximate abilities in human adults that even comes, surprisingly, from tasks that *do* ask for judgments based on Arabic numerals.

Brain imaging and lesion studies have shown that a localized area in the inferior parietal cortex may be specialized for the processing of quantities. Patients with damage to this area may lose the ability to perform very simple numerical comparisons or interval bisections (deciding what number comes between 5 and 7, for example). Such damage can leave performance on similar but non-numerical bisection tasks unaffected; for example, patients may be able to decide which letter comes between E and G with no trouble (Dehaene and Cohen 1997). (Inferior parietal lesions do not affect skills that involve numbers but depend on verbal rote learning, such as multiplication of single-digit numbers (Cohen, Dehaene, Chochon, Lehericy, and Naccache 2000). Brain imaging evidence supports the idea of an inferior parietal area involved in quantity manipulation. Tasks like numerical comparison, simple subtraction, interval bisection, and others that involve the processing of quantities tend to activate this area differentially, though rote number tasks, such as problems that call upon the memorized multiplication tables, do not (Dehaene, Tzourio, Frak, Raynaud, Chen, Mehler, and Mazoyer 1996; Dehaene, Dehaene-Lambertz, and Cohen 1998; Chochon, Cohen, van de Moortele, and Dehaene 1999). Both neuropsychology and imaging research have provided evidence that there is a functionally localizable representation of quantity in human adults, though it remains unknown how specifically this area is activated by number- and math-related tasks.

Infants and animals, at least under some circumstances, clearly are sensitive to the *numerosities* of stimulus sets; their behavior is based on number and not the non-numerical properties of the sets. It would be surprising if infants and animals possessed abstract nonlinguistic numerical abilities and adult humans did not. However, adults' patterns of performance in approximate number studies have been examined more finely than have animals' or infants', and it seems that judgments of approximate numerosity in humans are consistently found to be highly influenced by sensory properties of the stimulus, such as regularity in a visual array, or frequency in an auditory sequence (Massaro 1976; Ginsburg and Nicholls 1988; Ginsburg and Pringle 1988; Ginsburg 1991). Computational modeling of human numerosity judgments has also cast doubt on numerosity perception as an abstract, amodal process. Proposed models of visual numerosity estimation predict human performance quite accurately, though these models perform their estimations based on stimulus properties such as area, which are correlated with, but not equivalent to, stimulus numerosity (Allik and Tuulmets 1991; Allik, Tuulmets and Vos 1991; Allik and Tuulmets 1993). Because of these perceptual influences, "numerosity" perception has often been explained in terms of modality-specific perceptual processes (e.g. texture perception for visual arrays), or in terms of processes specific to stimuli in certain formats (e.g. timing mechanisms for temporally distributed elements). These explanations have been especially favored by the many researchers approaching the problem from the vantage point of the perceptual scientist (perhaps because these scientists may not be particularly interested in demonstrating the ability to abstract across modality).

If adults do have an abstract nonlinguistic sense of number, why should their numerical judgments show such stimulus-dependence? There are several possibilities. It may be that adults tend *not* to use number while performing approximate tasks; they may use continuous quantities such as size, area, duration, or density instead, depending on the type of stimulus in question. This does not necessarily mean that they are not capable of using number (which seems unlikely given the animal studies discussed earlier); it may simply mean that humans adopt strategies that deviate from the instructions they're given. Another alternative is that the subjects *do* use number to make their judgments, and there *is* an abstract number representation, but the perceptual properties of stimuli to be enumerated are influencing whatever process builds that representation. Until this issue of perceptual vs. numerical explanation is addressed and reconciled with the comparative studies, there will be a pronounced gap in our description of human numerical competence. This question will be addressed in the Chapter 2 of the thesis.

## 1.2 What is the nature of the representation that underlies this number concept?

Though each piece of experimentation regarding the abstract nature of numerical representation is not necessarily conclusive on its own, there is certainly reason to think that such a representation is likely to exist. How can abstract number be represented, then, without recourse to symbolic notation? It is easy to imagine representing a number of objects through a visuospatial representation, such as an area, or a number of sounds by remembering the length of the time interval they filled. But for the representation to be abstract, it must span different sensory modalities. Converging studies in many different populations suggest that number is represented spatially through analog magnitudes. These are described metaphorically as number lines or vessels filled with liquid; the essence of the idea is that discrete number is represented through a rough continuous quantity. These "mental magnitudes" are noisy: the representation is less precise for larger numbers and more precise for smaller numbers. To understand the basic properties of this representation through one useful analogy, picture a graduated cylinder filled with liquid, up to the value "x." The level of the liquid is not exactly known due to the noise in the representation, so it could be slightly more or less than four. Now, to represent "2x" instead of "x" the amount of liquid would roughly double, but so would the variability in the magnitude representation. This property can account for the fact that Weber's Law governs operations that use the magnitude representation. There is still active controversy about the true source of this Weber susceptibility (Dehaene 1992; Gallistel and Gelman 1992; Dehaene 1997; Whalen, Gallistel and Gelman 1999; Brannon, Wusthoff, Gallistel, and Gibbon 2001); indeed, there is controversy about whether the question can even be determined experimentally (Dehaene 2001). However, it is generally agreed that analog magnitude representations, common to humans and nonverbal animals, can account for a large body of experimental data on numerical competence.

These noisy magnitude representations, however, may not be able to account for all of the relevant data, including some of the results discussed earlier. Some experiments suggest that the nature of the number representation may depend on the range of numerosities in question. There is evidence that separable systems govern numerical abilities with small and large numbers. Because controversy remains regarding the existence of distinct cognitive systems that deal with smaller vs. larger numerosities, bodies of research which may be probing very different processes are sometimes discussed under the assumption that they form a unitary whole. The source of the two-systems theory is as follows. Early research on number processing in humans found that when judging numbers of dots, adult produced very typical patterns of performance. With very small numbers like 2, 3, or 4, reaction times were very fast and error rates were very low. Then at about 5 items, subjects' reaction times appeared to increase with each additional item, suggesting a counting process. Then for larger numbers of items, error rates increased a great deal and reaction times did not suggest counting. This was explained in terms of three separate processes: "estimation" for the largest numbers, "counting" for the middle range, and "subitization" for the very small numerosities (Kaufman, Lord, Reese, and Volkmann 1949). Subitization has been described as the "immediate perception" of a very small numerosity.

Subsequent studies have suggested that there is not actually a separate process that enumerates very small quantities, but this remains undecided (Mandler and Shebo 1982; Gallistel 1990; Balakrishnan and Ashby 1991; Balakrishnan and Ashby 1992; Dehaene 1992; Carey 2001). Infant work, on the other hand, does seem to be decidedly inconsistent with the idea of a single system that is subject to Weber's Law. If infants' numerical capabilities spring from analog magnitudes, then discriminability should be dependent only on the ratio of the numerosities to be discriminated. Yet infants can discriminate 2 from 3 dots but not 8 from 12 (Starkey and Cooper 1980; Xu and Spelke 2000). Also, they succeed at a "1 + 1 = 2" addition task but not a "5 + 5 = 10" addition task (Chiang and Wynn 2000), which has been presented as evidence that infants may not represent larger numerosities[1]. It has been suggested that these tasks that involve very small numbers of objects may be solved through infants' use of an attentional system that tracks a few objects at a time (Dehaene 1997; Carey 2001). This "object file" system may be responsible for the results seen with animal small number tasks as well. Therefore, when attempting to characterize the kinds of potentially biologically determined numerical representations that are common to humans and animals, that are subject to Weber's Law, and that may underlie symbolic operations, it is wise to be aware that data from tests with small numbers may not be tapping the desired processes. To ensure that analog magnitudes and not object files are being recruited, larger numerosities are necessary.

---

[1] However, just as stimulus properties may affect early stages of the enumeration process in adults without threatening the idea that they represent number, as discussed earlier, it is possible that a parallel argument applies to these infant addition studies. The problem may lie in the stimuli used in the 5 + 5 task, which because of their greater numerosities are very different from those in the 1 + 1 task, and may be more difficult for the infants to individuate or enumerate in the first place.

## 1.3 How to construct an analog magnitude

The proposed mechanisms for the construction of analog representations of magnitude may be divided into two broad classes: those that operate iteratively (such as a preverbal counting-like process (Gallistel and Gelman 1992; Whalen, Gallistel and Gelman 1999), and those that are non-iterative (for example, sampling approximate density and area, distance between elements, or rate and duration (Church and Broadbent 1990; Church and Broadbent 1991)). A long tradition of evidence from numerosity estimation tasks has led some researchers to posit iterative enumeration mechanisms such as preverbal counting (Gallistel and Gelman 1992), or protocounting (Davis and Perusse 1988). When adults are shown an array of dots, for example, and asked to make a speeded judgment of how many there are (by producing a number), response time reliably increases with the number of dots; this has been explained in terms of serial enumeration mechanisms. However, as the number of dots increases, our rough approximation of their numerosity becomes even rougher, and our fuzzy representation of the number of dots maps onto a larger set of possible symbolic responses. This increase in the number of response options could make response selection more difficult, accounting for the observed increases in reaction time.

As discussed earlier, numerosity judgments are extremely sensitive to perceptual properties of the stimuli to be enumerated. It is certainly possible for an iterative counting-like process to be affected by such properties. For example, any counting mechanism must individuate elements, and the effectiveness of this individuation process could certainly be influenced by changes in the arrangement or size of the elements. However the stimulus-dependence that has been observed in numerosity judgment can be more parsimoniously attributed to the operation of a non-iterative enumeration process, which uses multiple stimulus attributes in combination as cues to numerosity. Thus, information about a visual array such as its density and area might be transformed into a representation of the array's numerosity that is modality-independent. Under such conditions, any attribute of the stimulus that affects our perception of these cues (e.g. density aftereffects or anchoring effects) will clearly alter our representations of numerosity as well. The nature of this transformation from stimulus-specific properties to numerosity representations remains unknown; the same mechanism may serve to convert all spatially presented stimuli to abstract form, while another may perform the same task for all temporal presentations. The latter sort of mechanism may be responsible for animals' representation of the numerosity of a sequence of events; it has been proposed that animals keep track of the average interval between events and the overall sequence duration, using these two durations to compute the total number of events (Church and Broadbent 1990; Church and Broadbent 1991).

A decisive piece of experimental support for one or the other of these models – iterative or non-iterative – would be the first step in clarifying the nature of the processes that lead to the construction of analog representations of number. This could lead to further investigation of the generality of these processes: how are sequential sets enumerated compared to simultaneously presented sets? Are sets presented to different sensory modalities enumerated differently? The final experiment of Chapter 2 attempts to make the initial distinction between iterative and non-iterative models of enumeration.

### 1.4 Do analog magnitudes enter into arithmetical computations?

Evidence certainly exists that animals, human infants, and human adults can alter their behavior based on number, and that these representations of number may be truly abstract. Are these representations functional in the sense that they are available to do conceptual work in arithmetic operations? The depth of the preverbal number concept in humans and animals is unknown. If it is true that the preverbal number sense is a biologically based stepping-stone supporting the symbolic numerical abilities developed by humans, we might expect to see the beginnings of arithmetical manipulation of quantity even in other species.

Studies with various nonhuman animals demonstrate remarkable arithmetic skills even though the subjects cannot make use of a symbolic number system (Church and Meck 1984). Though most of these studies involve extensive training, untrained wild rhesus monkeys have been found capable of the same simple addition and subtraction operations that Wynn demonstrated in human infants, and more (Hauser, MacNeilage, and Ware 1996; Sulkowski and Hauser 2001). While alternative perceptual explanations have not been entirely ruled out, it seems that there is promising evidence for arithmetic capabilities in some primates. Human infants also can perform simple arithmetic operations such as addition and subtraction at least on small numbers of objects. When presented with displays in which 2 items are placed behind a screen and then one is removed, infants expect exactly one item to remain. Similarly, when two items are sequentially placed behind a screen, infants expect there to be exactly two (Wynn, 1992).

These particular infant and monkey studies, like many others, involved the manipulation of exact very small numerical quantities. As discussed earlier, it is very possible that processing of such quantities may be distinct from that of larger approximate numerosities (Dehaene 1997; Carey 2001; but see Gallistel 1990). Therefore these studies may not deal with mechanisms analogous to those that could allow human adults to perform approximate arithmetical operations on large sets. While infants can discriminate between larger sets as well, given large enough comparison ratios (Xu and Spelke 2000), it is not clear whether these ranges of numerosities may also be used in arithmetic operations. Some recent data suggest that they cannot (Chiang and Wynn 2000). However, there is extensive evidence that animals process representations of larger numbers arithmetically. Research on foraging behavior has shown that foraging patterns seem to depend on animals' calculation of rate of return, which requires a complex combination of duration and number information (Leon and Gallistel 1998). A large body of research on temporal processing, which is likely to depend upon representations very similar to representations of numerical quantity, has demonstrated arithmetic processing of durations (Meck, Church, and Gibbon 1985; Gibbon, Malapani, Dale, and Gallistel; 1997).

More direct assessments of arithmetic capabilities with larger quantities have also produced convincing results, many of which suggest that certain arithmetic operations such as summation may be automatic when two quantities are presented closely in space and/or time. When rats see two lights and hear two sounds at once, they give the conditioned response that means "4" (Church and Meck 1984). Spontaneous summation has also been demonstrated in many other species; chimpanzees have been found to sum quantities presented both nonsymbolically (without training) and symbolically (with training of the symbolic associations, but no training of the summation process) (Rumbaugh, Savage-Rumbaugh, and Hegel 1987; Rumbaugh, Savage-Rumbaugh, and Pate 1988; Boysen and Berntson 1989). Chimpanzees are not the only animals capable of symbolic operations like addition. Spontaneous summation of trained symbols has also been demonstrated by Olthof and colleagues in squirrel monkeys and in pigeons; however, the data suggest that the quantities being summed were masses rather than numerosities (Olthof, Iden, and Roberts 1997; Olthof and Roberts 2000).

Human adults tend to sum small animals do with nonsymbolic quantities (LeFevre, Bisanz, and Mrkonjic 1988; Stadler, Geary, and Hogan 2001). We do not know whether human adults can perform arithmetical operations when they are prevented from using their learned symbolic abilities. However, just as symbolic number comparisons draw upon analog magnitude representations (Dehaene, Dupoux, and Mehler 1990), fMRI studies have shown that symbolic arithmetic can recruit those same representations. When simple addition is performed on single Arabic digits, the areas that show activation are associated with rote learning of number facts. When the subject must choose an answer that is only approximately correct, however, so that the rote-learning method will not give successful responses, areas associated with analog magnitude processing are activated (Dehaene, Spelke, Pinel, Stanescu, and Tsivkin 1999). The question of *nonsymbolic* arithmetical capabilities in humans remains untouched, though it plays a key role in the larger issue of preverbal number's importance to human mathematical development.

This thesis addresses some of the remaining questions about the preverbal number sense in human adults regarding the degree to which number representations are truly abstract, the possible means by which we construct these representations, and the roles they play in calculation processes. Based on research on numerical cognition in nonhuman animals, human infants, and human adults so far, it is plausible to suppose that the representational system underlying basic numerical competence is common to many species. It is a system available to nonverbal animals and prelinguistic infants, and it may well have provided humans with the necessary foundation for the development of the intricate mathematical structures they have constructed. This set of studies extends existing research, deepening our understanding of the number concept in human adults, its relationship to the number sense of preverbal beings, and its potential contributions to the development of symbolic arithmetic.

## References


Allik, J. and T. Tuulmets (1991). "Occupancy model of perceived numerosity." Perception & Psychophysics **49**(4): 303-314.

Allik, J. and T. Tuulmets (1993). "Perceived numerosity of spatiotemporal events." Perception & Psychophysics **53**(4): 450-9.

Allik, J., T. Tuulmets, and P. G. Vos (1991). "Size invariance in visual number discrimination." Psychological Research **53**(4): 290-295.

Antell, S. and D. Keating (1983). "Perception of numerical invariance in neonates." Child Development **54**: 695-701.

Balakrishnan, J. D. and F. G. Ashby (1991). "Is subitizing a unique numerical ability?" Perception & Psychophysics **50**(6): 555-564.

Balakrishnan, J. D. and F. G. Ashby (1992). "Subitizing: Magical numbers or mere superstition?" Psychological Research **54**(2): 80-90.

Bijeljac-Babic, R., J. Bertoncini, and J. Mehler (1991). "How do four-day-old-infants categorize multi-syllabic utterances?" Developmental Psychology **29**: 711-721.

Boysen, S. T. and G. G. Berntson (1989). "Numerical competence in a chimpanzee (Pan troglodytes)." J Comp Psychol **103**(1): 23-31.

Brannon, E. M., C. J. Wusthoff, C. R. Gallistel, and J. Gibbon (2001). "Numerical subtraction in the pigeon: evidence for a linear subjective number scale." Psychol Sci **12**(3): 238-43.

Carey, S. (2001). Bridging the gap between cognition and developmental neuroscience: the example of number representation. Handbook of Developmental Cognitive Neuroscience. C. A. Nelson and M. Luciana, MIT Press: 415-431.

Chiang, W. and K. Wynn (2000). "Infants' representation and tracking of objects: implications from collections." Cognition **77**: 169-195.

Chochon, F., L. Cohen, P. F. van de Moortele, and S. Dehaene (1999). "Differential contributions of the left and right inferior parietal lobules to number processing." J Cogn Neurosci **11**(6): 617-30.

Church, R. M. and H. A. Broadbent (1990). "Alternative representations of time, number, and rate." Cognition **37**(1-2): 55-81.

Church, R. M. and H. A. Broadbent (1991). A connectionist model of timing. Commons, Michael L. (Ed); Neural network models of conditioning and action. Quantitative analyses of behavior series. (pp. 225 240). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

Church, R. M. and W. H. Meck (1984). The numerical attribute of stimuli. Animal Cognition. H. L. Roitblatt, T. G. Bever and H. S. Terrace. Hillsdale, NJ, Erlbaum: 445-464.

Cohen, L., S. Dehaene, F. Chochon, S. Lehericy, and L. Naccache (2000). "Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study." Neuropsychologia **38**(10): 1426-40.

Davis, H. and R. Perusse (1988). "Numerical competence in animals: Definitional issues, current evidence, and a new research agenda." Behavioral and Brain Sciences **11**(4): 561-615.

Dehaene (1997). The Number Sense. New York, Oxford University Press.

Dehaene, S. (1992). "Varieties of numerical abilities. Special Issue: Numerical cognition." Cognition **44**(1-2): 1-42.

Dehaene, S. (2001). "Subtracting pigeons: logarithmic or linear?" Psychol Sci **12**(3): 244-6; discussion 247.

Dehaene, S. and L. Cohen (1997). "Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic." Cortex **33**(2): 219-50.

Dehaene, S., G. Dehaene-Lambertz, and L. Cohen (1998). "Abstract representations of numbers in the animal and human brain." Trends Neurosci **21**(8): 355-61.

Dehaene, S., E. Dupoux, and J. Mehler (1990). "Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison." Journal of Experimental Psychology: Human Perception & Performance **16**(3): 626-641.

Dehaene, S., E. Spelke, P. Pinel, R. Stanescu, and S. Tsivkin (1999). "Sources of mathematical thinking: behavioral and brain-imaging evidence [see comments]." Science **284**(5416): 970-4.

Dehaene, S., N. Tzourio, V. Frak., L. Raynaud, L. Cohen, J. Mehler, and B. Mazoyer (1996). "Cerebral activations during number multiplication and comparison: a PET study." Neuropsychologia **34**(11): 1097-106.

Gallistel, C. R. (1990). The organization of learning. Cambridge, MA, US, MIT Press.

Gallistel, C. R. and R. Gelman (1992). "Preverbal and verbal counting and computation. Special Issue: Numerical cognition." Cognition **44**(1-2): 43-74.

Gibbon, J., C. Malapani, C. L. Dale, and C. R. Gallistel 1997). "Toward a neurobiology of temporal cognition: advances and challenges." Curr Opin Neurobiol **7**(2): 170-84.

Ginsburg, N. (1991). "Numerosity estimation as a function of stimulus organization." Perception **20**(5): 681-686.

Ginsburg, N. and A. Nicholls (1988). "Perceived numerosity as a function of item size." Perceptual & Motor Skills **67**(2): 656-658.

Ginsburg, N. and L. Pringle (1988). "Haptic numerosity perception: Effect of item arrangement." American Journal of Psychology **101**(1): 131-133.

Hauser, M. D., P. MacNeilage, and M. Ware (1996). "Numerical representations in primates." Proc Natl Acad Sci U S A **93**(4): 1514-7.

Kaufman, E. L., M. W. Lord, T.W. Reese, and J. Volkmann (1949). "The discrimination of visual number." American Journal of Psychology **62**: 498-525.

Koechlin, E., S. Dehaene, and J. Mehler (1998). "Numerical transformations in five-month-old human infants." Mathematical Cognition **3**: 89-104.

LeFevre, J. A., J. Bisanz, and L. Mrkonjic (1988). "Cognitive arithmetic: evidence for obligatory activation of arithmetic facts." Mem Cognit **16**(1): 45-53.

Leon, M. I., and C. R. Gallistel (1998). Self-stimulating rats combine subjective reward magnitude and subjective reward rate multiplicatively. Journal of Experimental Psychology: Animal Behavior Processes, 24(3), 265-277.

Lipton, J. and E. Spelke (2000). Infants' discrimination of large numbers of auditory sequences. Poster presented at MIT Cognitive Science Forum, March.

Mandler, G. and B. J. Shebo (1982). "Subitizing: an analysis of its component processes." J Exp Psychol Gen **111**(1): 1-22.

Massaro, D. W. (1976). "Perceiving and counting sounds." J Exp Psychol [Hum Percept] **2**(3): 337-46.

Mechner, F. (1958). "Probability relations within response sequences under ratio reinforcement." Journal of the Experimental Analysis of Behavior **1**: 109-122.

Mechner, F. and L. Guevrekian (1962). "Effects of deprivation upon counting and timing processes." Journal of the Experimental Analysis of Behavior **5**: 463-466.

Meck, W. H. and R. M. Church (1983). "A mode control model of counting and timing processes." Journal of Experimental Psychology: Animal Behavior Processes **9**(3): 320-334.

Meck, W. H., R. M. Church, and J. Gibbon (1985). "Temporal integration in duration and number discrimination." Journal of Experimental Psychology: Animal Behavior Processes **11**(4): 591-597.

Mix, K. S., J. Huttenlocher, and S. C. Levine (1996). "Do preschool children recognize auditory-visual numerical correspondences?" Child Dev **67**(4): 1592-608.

Mix, K. S., S. C. Levine, and J. Huttenlocher (1997). "Numerical abstraction in infants: another look." Dev Psychol **33**(3): 423-8.

Moore, D., J. Benenson, J. S. Reznick, and M. Peterson (1987). "Effect of auditory numerical information on infants' looking behavior: Contradictory evidence." Developmental Psychology **23**(5): 665-670.

Moyer, R. S. and T. K. Landauer (1967). "Time required for judgements of numerical inequality." Nature **215**(109): 1519-20.

Olthof, A., C. M. Iden, and W. A. Roberts (1997). "Judgements of ordinality and summation of number symbols by squirrel monkeys (Saimiri sciureus)." J Exp Psychol Anim Behav Process **23**(3): 325-39.

Olthof, A. and W. A. Roberts (2000). "Summation of symbols by pigeons (Columba livia): the importance of number and mass of reward items." J Comp Psychol **114**(2): 158-66.

Platt, J. R. and D. M. Johnson (1971). "Localization of position within a homogeneous behavior chain: effects of error contigencies." Learning and Motivation **2**: 386-414.

Rumbaugh, D. M., E. S. Savage-Rumbaugh, and J. L. Pate (1988). "Addendum to "Summation in the chimpanzee (Pan troglodytes)"." J Exp Psychol Anim Behav Process **14**(1): 118-20.

Rumbaugh, D. M., S. Savage-Rumbaugh, and M. T. Hegel (1987). "Summation in the chimpanzee (Pan troglodytes)." J Exp Psychol Anim Behav Process **13**(2): 107-15.

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

Stadler, M., D. Geary, and M. Hogan (2001). "Negative priming from activation of counting and addition knowledge." Psychological Research **65**: 24-27.

Starkey, P. and R. Cooper (1980). "Perception of numbers by human infants." Science **210**: 1033-1035.

Starkey, P., E. S. Spelke, and R. Gelman (1983). "Detection of intermodal numerical correspondences by human infants." Science **222**(4620): 179-181.

Starkey, P., E. S. Spelke, and R. Gelman (1990). "Numerical abstraction by human infants." Cognition **36**(2): 97-127.

Sulkowski, G. M. and M. D. Hauser (2001). "Can rhesus monkeys spontaneously subtract?" Cognition **79**(3): 239-62.

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

Whalen, J., C. R. Gallistel, and R. Gelman (1999). "Nonverbal counting in humans: The psychophysics of number representation." Psychological Science **10**(2): 130-137.

Wynn, K. "Psychological foundations of number: numerical competence in human infants." Trends in Cognitive Science.

Wynn, K. (1992). "Addition and subtraction by human infants." Nature **358**: 749-750.

Wynn, K. (1996). "Infants' individuation and enumeration of physical actions." Psychological Science **7**: 164-169.

Xu, F. and E. S. Spelke (2000). "Large number discrimination in 6-month-old infants." Cognition **74**(1): B1-B11.

# Chapter 2

# The Construction of Large Number Representations in Adults


Abstract

What is the nature of our mental representation of quantity? We find that human adults show no performance cost of comparing numerosities across versus within visual and auditory stimulus sets, or across versus within simultaneous and sequential sets. In addition, reaction time and performance in such tasks are determined by the ratio of the numerosities to be compared; absolute set size has no effect. These findings suggest that modality-specific stimulus properties undergo a non-iterative transformation into representations of quantity that are independent of the modality or format of the stimulus.

## 2.0 Introduction

Substantial experimental evidence points to the idea that humans possess an abstract sense of approximate quantity, or "number sense." Converging evidence from studies of numerical competence in normal adults, patients, infants, young children, and nonhuman animals has led many investigators to conclude that a domain-specific system of knowledge, present in many species, is responsible for the sense of number and forms the basis for the complex symbolic manipulation of number developed by humans (e.g. Dehaene, 1997; Gallistel & Gelman, 1992). Many questions remain, however, regarding the nature of number representation and the processes that construct it.

A truly abstract number sense would be capable of representing the numerosity of any set of discrete elements, whether events or objects, homogeneous or heterogeneous, or simultaneous or sequential. Nonhuman animals and human infants have been found to generalize across stimuli on the basis of numerosity, independent of stimulus shape, color, or identity; these results have been taken as evidence of abstract numerical ability (Gallistel and Gelman 1992; Dehaene, Dehaene-Lambertz and Cohen, 1998). Human adults clearly have this ability when dealing with exact quantities labeled with Arabic numerals. Yet judgments of approximate numerosity in humans are consistently found to be highly influenced by sensory properties of the stimulus, such as regularity in a visual array, or frequency in an auditory sequence (Massaro 1976; Ginsburg and Nicholls 1988; Ginsburg and Pringle 1988; Ginsburg 1991). For example, the density and grouping patterns of large visual arrays affect the numerosity judgments of human adults (Durgin 1995). Because of these perceptual influences, "numerosity" perception is often explained in terms of modality-specific perceptual processes (e.g. texture perception for visual arrays), or in terms of processes specific to stimuli in certain formats (e.g. timing mechanisms for temporally distributed elements). Computational modeling of human numerosity judgments has also cast doubt on numerosity perception as an abstract, amodal process. Proposed models of visual numerosity estimation predict human performance quite accurately, though these models perform their estimations based on stimulus properties such as area, which are correlated with, but not equivalent to, stimulus numerosity (Allik and Tuulmets 1991; Allik, Tuulmets, and Vos 1991; Allik and Tuulmets 1993). A thorough explanation of human numerical competence must account for this stimulus-dependence.

Research on numerical competence in human infants has produced its own inconsistencies. The ability to relate numerosities of sets presented to different sensory modalities is a requirement of any system that can be said to represent abstract numerosity. Looking times in six- to eight-month-old infants have been found to depend on the correspondence between the number of objects in a visual array and the number of drumbeats in a sequence (Starkey, Spelke, and Gelman 1983; Moore, Benenson, Reznick, and Peterson 1987; Starkey, Spelke, and Gelman 1990). However, a recent replication attempt found no preference in either direction (Mix, Levine, and Huttenlocher 1997). In addition, 3-year-old children do not perform well on crossmodal numerosity matching tests, in which they must choose the visual-spatial array that corresponds in number to a sequence of sounds (Mix, Huttenlocher, and Levine 1996). While it is certainly possible that infants possess modality-independent numerical ability, to date evidence of this ability is equivocal.

The animal literature provides far more robust evidence of the use of modality-independent numerical information. An impressive example of crossmodal transfer has been demonstrated in rats (Church and Meck 1984). The animals were trained to press one lever when presented with two lights or white noise bursts, and another lever when presented with four lights or white noise bursts. When two lights and two sounds were presented together, the rats pressed the "four" lever, suggesting that they spontaneously combined the quantities of light and sound and responded to their sum. This occurred even though the compound stimulus was a combination of two other stimuli, each of which taken alone demanded a different response.

When a previously unseen stimulus was used (one sound and one light) the rats again responded to the sum, pressing the "two" lever.

Thus, animal research shows that rats are capable of crossmodal numerosity combination; infant research does not provide such a clear picture. In human adults, recent studies have shown that a numerosity perceived through the presentation of an auditory stimulus influences the perceived numerosity of a simultaneous visual stimulus, at least for very small numerosities (Shams, Kamitani, and Shimojo 2000). Human adults, of course, are able to make crossmodal comparisons of all sizes through the use of the symbolic number system. When adults are prevented from using this symbolic system, how do they perform in crossmodal comparison tasks? The answer to this question should depend upon the specific mechanisms used for each numerosity judgment. The present study examined performance in crossmodal numerical comparison tasks, in an effort to shed light on the nature of the processes and representations involved in judgments of relative numerousness. Whether these comparison experiments can speak to the processes and representations involved in judgments of *absolute* numerosity as well is a matter of debate. It is possible that such judgments require very different forms of numerical competence, and that animals can only deal with relative numerosities, unlike humans who may represent absolute numerosities as well (Davis and Perusse 1988). Therefore the present study may not generalize to the entire range of human estimation abilities, but it has the advantage of being relevant to the type of numerical competence that is likely to apply across species.

The use of a crossmodal comparison task also allows us to confront an important question regarding the usefulness of numerical comparison tasks in general. Researchers have used tasks involving numerical comparisons repeatedly to draw conclusions about how we process number, but serious confounds may be introduced by these tasks. Participants may in many cases be able to complete the task without using numerosity as the actual basis for comparison. This problem could exist even for studies in which non-numerical stimulus properties such as element density are carefully controlled. To understand how these tasks might be carried out, imagine an observer faced with two briefly presented arrays of dots. In order to make a comparative judgment, the observer could enumerate the dots in each array (by some iterative counting-like process, or perhaps by combining information about the total area of the display and its average density) and base his/her judgment on a truly numerical representation. Most numerical comparison tasks depend on the assumption that this is indeed an accurate description of the process in question. Alternatively, however, the task could instead be performed by the use of an intermediate perceptual representation that is neither numerical nor amodal. For example, the comparison of two arrays could be based on perceptual representations of the arrays (including their areas and/or densities), rather than on more abstract representations of the numerical quantities in each. Such a perceptual representation could be specific to visual processing, and it would contain numerosity information only implicitly. This scenario is depicted in Figure 1a. Both dot arrays are presented visually, and the assumption is that the observer will compare them based on an amodal quantity representation (the graduated cylinders on the right). However, it is possible that an intermediate modality-specific representation of quantity exists as well (depicted by the question mark). If there is such a modality-specific representation, any task involving a comparison of two visual arrays could be accomplished without calling upon an amodal representation.

Here, we attempt to address these concerns experimentally. By using a crossmodal numerical comparison task, we ensure that participants cannot succeed by comparing intermediate perceptual representations as described above (see Fig. 1b). If numerosity judgments are made on the basis of non-numerical information, then crossmodal judgments should show performance deficits relative to intramodal judgments. However, if a truly abstract representation of quantity exists, then we might find little or no cost for crossmodal comparisons relative to unimodal comparisons. In the present studies, we use comparisons of large approximate numerosities to determine the relative difficulties of crossmodal and intramodal numerosity judgments. This allows us both to explore adults' ability to manipulate numerical quantities presented in different modalities, and, in the process, to determine the usefulness of comparison tasks to research on numerical competence. We use two different complementary numerical comparison judgments to ensure that our results generalize across tasks. We then extend this crossmodal comparison method further in order to distinguish between proposed mechanisms for the construction of numerosity representations.

## 2.1 Chapter 2 Experiment 1

Experiment 1 investigated whether adults can compare numerosities crossmodally as accurately as they can intramodally. Participants made numerosity judgments about stimulus pairs that consisted of two sequences of flashes ("Visual"), two tone sequences ("Auditory"), or a flash sequence and a tone sequence ("Crossmodal").

### *Method*

Participants. Five males and ten females between the ages of 18 and 35 participated in the study. All had normal or corrected-to-normal hearing and vision and were paid $8 for their participation.

Apparatus. Participants sat in a small darkened room at a distance of approximately 60 cm from the presentation screen. Visual stimuli were presented on a Sony Multiscan monitor by a Power Macintosh 8600 computer. Auditory stimuli were presented from the Macintosh's built-in speaker. The apparatus was the same for all of the experiments that follow.

Design. All participants received the same stimulus conditions in counterbalanced order. Auditory, Visual, and Crossmodal trials were blocked; participants received two blocks of 24 trials of each condition, presented in ABCCBA order, for a total of 48 trials per condition. Before the experimental blocks began, there were 10 practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block (Auditory, Visual, or Crossmodal). The stimulus pairs presented were 10-10, 20-20, 30-30, 10-20, 10-30, and 20-30. "Same" and "Different" trials were equally frequent.

Stimuli. An "Auditory" sequence consisted of a series of 10, 20, or 30 tones. All tones in a particular sequence had the same duration, but tone duration varied from sequence to sequence between ~20 and ~60 ms, with the interval between tones remaining constant at ~50 ms. Tone presentation rate ranged from 7 to 11 tones per second, varying randomly so that the duration of the entire sequence was not a reliable cue to numerosity. The end of each tone sequence was marked by a brief high-pitched beep. A "Visual" trial consisted of a sequence of 10, 20, or 30 small white circles (diameter ~1 cm) appearing at the location of the fixation cross. The timing of the visual sequences was slightly slower than that of the auditory sequences: durations of the circles' flashes varied from ~30 ms to ~80 ms. Flash presentation rate ranged from 6 to 9 flashes per second, also varying randomly so that the duration of the entire sequence was not a reliable cue to numerosity..

Procedure. In a "Visual" trial, a small red fixation cross appeared for ~400 ms, followed by a flash sequence, a pause of ~930 ms, and a second flash sequence. Responses were made after the second flash sequence ended. Participants were instructed to press one button if the number of flashes in the first sequence was the same as the number in the second sequence, and a second button if the numbers were different. In an "Auditory" trial, a sequence of tones played, followed by a pause of ~930 ms and a second tone sequence. In half of the "Crossmodal" trials, the visual sequence was presented first, and in the other half the auditory sequences came first; these trial types were interleaved within the Crossmodal block.

Measures. We focus on subjects' accuracy and not reaction time. All of these experiments involve temporal sequences, which make RT data difficult to interpret. For example, in many of the trials, the first sequence will be much less numerous than the second. The subject may well decide upon an answer of "different" long before the second sequence ends and the response interval begins. This issue applies to any conditions involving stimulus presentations that are temporally extended.

If subjects showed perfect cross-modal transfer of numerical information across modalities, then their performance on the cross-modal trials should be limited only by their ability to detect numerosity in each of the individual modalities. Because individual subjects might vary in their abilities to represent numerosity in visual vs. auditory temporal arrays, we determined separately for each subject which of the two intramodal numerical comparison tasks was more difficult for them (the "worse unimodal condition") and compared performance across subjects in this condition to performance in the cross-modal condition.

Results.

Figure 2 shows mean accuracy for each condition in the left panel. The "worse unimodal" vs. Crossmodal comparison (see Measures) is shown in the right hand panel of Figure 2. The "worse unimodal" and Crossmodal means were both 73%; these were clearly not significantly different ($t(14) = -0.03$, $p > 0.9$). Figure 2a depicts the way accuracy varied as a function of the size of the difference in the two sets to be discriminated. Only the "different" trials are shown in this plot, because the "different" trials may be clearly categorized by ratio (10:30, 10:20, 20:30), but the "same" trials may not (10:10, 20:20, 30:30). An ANOVA (Condition x Ratio) revealed a significant effect of Ratio ($p < .0005$); there was no effect of Condition.

Discussion.

We found no performance cost for the Crossmodal comparison task compared to the Worse Unimodal comparison. In this experiment, the visual sequence comparisons were worse than the auditory sequence comparisons for 11 of the 15 participants; this is not surprising considering the well-known superiority of auditory processing over visual when presentation is temporal[2] (Lechelt 1975). The remarkably similar performance patterns found in the Worse Unimodal and Crossmodal conditions show that subjects have no trouble comparing numerosities across modalities. Modality-specific numerosity representations could not have been used to accomplish this task. However, the numerosities of items in a temporal sequence may well be enumerated and/or represented differently from the numerosities of items in a spatial and simultaneous array. The task used in Exp. 1 therefore could have allowed subjects to use a numerosity representation that was abstract in the sense that it was not modality-specific, while failing to be abstract in terms of the temporal vs. spatial format of the stimulus. Experiment 2 tested this possibility by requiring subjects to make similar comparisons within a single modality but across stimulus presentation formats.

## 2.2 Chapter 2 Experiment 2

In Experiment 2, subjects compared numerosities across visual arrays that differed in format rather than modality: temporal sequences of light flashes vs. simultaneous spatial arrays

[2] In the current studies, in each experiment involving auditory sequences, participants who were experienced musicians outperformed nonmusicians. The single participant who performed perfectly for the Auditory trials in Experiment 2 explained after her session that she was a tympani player, that the rate of tone presentation was within the range of the speed of a drum roll, and that she was used to both hearing and producing such rapid sequences.

of dots. Each subject therefore made numerical comparisons under three conditions: "Spatial," "Temporal," and "Cross-format."

*Method*

Participants. Eight males and six females between the ages of 18 and 35 participated in the study. All had normal or corrected-to-normal hearing and vision. One additional male participated in the study, but his data were excluded from further analysis due to self-reported noncompliance with experimenters' instructions.

Design. All participants received the following stimulus conditions in counterbalanced order. Spatial, Temporal, and Cross-Format trials were blocked; participants received two blocks of 48 trials of each condition, for a total of 96 trials per condition. Before the experimental blocks began, there were 20 practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block.

Stimuli. Experiment 2 used the visual sequence trials from Experiment 1 (now termed "Temporal"). In the "Spatial" trials, participants were presented with pairs of visual arrays. A visual array consisted of 10, 20, or 30 small black dots on a mid-gray background. The arrays of dots were presented inside an imaginary square measuring ~13 by 13 cm. The distribution of the dots was pseudorandom, though they did not touch or overlap. All of the dots in a particular array were the same size, but dot diameter varied from array to array between 0.2 and 0.6 cm.

Procedure. The procedure was the same as that of Experiment 1, except for the replacement of the auditory sequences of Experiment 1 with visual arrays in Experiment 2. Because of the brevity of the visual array compared to the visual sequence, a delay was introduced in the Cross-Format trials between the two presentations in an attempt to equalize memory demands for Temporal and Cross-Format conditions. The delay had a pseudorandomly selected duration ranging from that of the shortest visual sequence (~400 ms) to that of the longest (~2500 ms). The second presentation appeared after this delay period.

Results.

The left-hand plot in Figure 3 shows mean accuracy for each condition. As in Experiment 1, we determined the worse unimodal case for each participant individually (auditory or visual) and compared the "worse unimodal" mean to the Crossmodal mean. This is shown in the right hand panel of Figure 2. The "worse unimodal" mean was 71% and the Crossmodal mean was 70%; these were not significantly different ($t(14) = 0.33$, $p > 0.7$). Figure 3a depicts the way accuracy varied as a function of the size of the difference in the two sets to be discriminated. As in the corresponding plot from the previous experiment, only the "different" trials are shown in this plot, because the "different" trials may be clearly categorized by ratio (10:30, 10:20, 20:30) but the "same" trials may not (10:10, 20:20, 30:30). An ANOVA (Condition x Ratio) revealed a significant effect of Ratio ($p < .0005$); there was no effect of Condition.

Discussion.

The results of Experiment 2 demonstrate that the comparison task was not performed using format-specific numerosity representations. Experiments 1 and 2 show that comparing approximate numerosities across different modalities or formats is no more difficult than comparing within modalities or formats; this strongly suggests that participants in these studies have formed true abstract representations of approximate numerosity and used these representations as the bases of their comparative judgments. If this is the case, participants should be able to compare numerosity across both format and modality at once; this task was performed in Experiment 3.

## 2.3 Chapter 2 Experiment 3

In this experiment, participants were asked to compare numerosities of spatially presented visual stimuli and temporally presented auditory stimuli. Participants made relative numerosity judgments about stimulus pairs that consisted of two dot arrays ("Visual/Spatial"), two tone sequences ("Auditory/Temporal"), or a dot array and a tone sequence ("Crossmodal/Cross-Format," shortened for ease to "Cross").

### Method.

Participants. Five males and nine females between the ages of 18 and 35 participated in the study. All had normal or corrected-to-normal hearing and vision. One additional female was excluded after falling asleep during the study.

Design. All participants received the following stimulus conditions in counterbalanced order. Auditory/Temporal, Visual/Spatial, and Cross trials were blocked; participants received two blocks of 48 trials of each condition, for a total of 96 trials per condition. Before the experimental blocks began, there were 20 practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block.

Stimuli. Visual arrays were the same as those used in Experiment 2; auditory sequences were the same as those used in Experiment 1.

Procedure. The procedure was again the same as that used in Experiment 2, except for the replacement of Experiment 2's visual sequences with visual arrays. A delay was introduced into the Cross trials as in Experiment 2.

Results.

Figure 4 shows mean accuracy for each condition. As in Experiments 1 and 2, we determined the worse unimodal/uniformat (abbreviated to "Worse Uni") case for each participant individually (auditory/temporal or visual/spatial) and compared the Worse Uni mean to the Cross mean. This is shown in the right hand panel of Figure 4. The Worse Uni mean was 77% and the Cross mean was 70%; these were significantly different ($t(13) = 2.44$, $p < 0.03$). Figure 4a depicts the way accuracy varied as a function of the size of the difference in the two sets to be discriminated. As in the corresponding plots from Experiments 1 and 2, only the "different" trials are shown in this plot. An ANOVA (Condition x Ratio) revealed a significant effect of Ratio ($p < .0005$); there was no effect of Condition.

In this experiment, large differences in reaction time patterns across subjects appeared to reflect the adoption of very different strategies. Some subjects reported performing the crossmodal task by the use of a 1:1 correspondence strategy; these people described matching the tones to the dots as the auditory sequence played[3]. When the tone sequence came first, they often reported "playing it back" and their response times reflected this strategy. Cross RTs for subjects who used this strategy were typically about twice as long when the auditory sequence came first than when the dot array came first. Participants who did not report using this strategy did not show such differences between the "auditory sequence first" and "visual array first" conditions.

Discussion.

Experiment 3 demonstrates that adults are able to compare numerosities across stimulus modality *and* format nearly as well as they can make comparisons within modality and format.

---

[4]This process was described by subjects as "imagining the tones painting the dots on a wall" or, notably, "imagining shooting a bullet at a dot each time I heard a tone." This last description led to the term "the bullethole strategy."

Because there is some difference between mean Worse Uni and Cross performance, there may be some cost for comparing numerosities in the Experiment 3 Cross condition. This pattern of results is somewhat different from those seen in Exps. 1 and 2. The crossmodal/cross-format cost suggests that when the numerosities to be compared are presented in different formats and modalities at the same time, comparison becomes more difficult. Why do participants have no trouble comparing across modality or format alone, while comparing across these two factors in combination causes errors? There are several possibilities. One is that this experiment encouraged subjects to use a disadvantageous strategy. In fact, use of the "bullethole" strategy, in which subjects attempted to use a 1:1 correspondence to complete the task, was reported much more frequently in the present experiment than in Experiment 2, which also used spatial and temporal stimuli. If this strategy is less effective than comparison based on abstract numerical representations, this could contribute to the deficit seen in Experiment 3.

Another possible explanation for the deficit lies within the nature of the task itself. Participants had to judge the stimulus numerosities as "same" or "different"; it may be that stimuli that are different in both modality and format are more likely to be judged "different" than stimuli that differ along only one of these dimensions. The data do provide some evidence for this hypothesis. If only the "different" trials are considered in this experiment, there is no crossmodal/cross-format cost – the performance cost is found only in the "same" trials. This suggests that participants could simply have been reluctant to judge an auditory sequential presentation the "same" as a visual simultaneous presentation. Though there was no main effect or interaction of Trial Type in our analysis, this conflict may contribute to the slight drop in performance across modalities and formats.

## 2.4 Chapter 2 Experiment 4

The first two experiments established that adults are able to compare numerosities crossmodally as easily as they perform comparisons intramodally, and that they are also able to compare easily across stimulus formats. When stimuli differ in both modality and format, as in Experiment 3, comparison may become more difficult. Experiment 4 explores the nature of the enumeration mechanism used in these comparison tasks. In addition, it tests one possible explanation for the greater difficulty observed within the crossmodal/cross-format condition of Experiment 3 and it deals with other objections that could be raised to the first three experiments.

The previous results suggest that abstract numerosity representations were the bases for participants' comparative judgments, raising the question of how these abstract representations are derived. The enumeration mechanisms that have been proposed may be divided into two broad classes: those that operate iteratively (such as a preverbal counting-like process (Gallistel and Gelman 1992)), and those that are non-iterative (for example, sampling approximate density and area, distance between elements, or rate and duration (Church and Broadbent 1990)). The present experiment was designed to distinguish between these classes of mechanisms. Like Experiment 3, this study used visual arrays and auditory sequences. We employed 5 different comparison ratios and 4 different absolute set sizes in order to assess the effects of set size and ratio on reaction time. Any iterative process of numerosity estimation should require more time for larger absolute set sizes. A non-iterative process, on the other hand, is less likely to require additional time for the enumeration of larger sets, so in this case only comparison ratios, and not set sizes, would determine reaction time.

The other changes in Exp. 4 were introduced to meet possible objections to aspects of Exps. 1-3. First, The use of only three numerosities in the previous studies could in principle make it possible for participants to perform the task by classifying each sequence or array as small, medium, or large, and comparing on the basis of these categories. While this strategy would indeed require a broad quantity-based judgment, it would be useful to observe performance on a more difficult task, which discourages classification strategies by employing more stimulus numerosities. Second, the comparison of the cross task to the worse uni task introduced the possibility that the costs of crossmodal or cross-format comparisons were artificially masked by regression effects, yielding an overestimate of the difficulty of the unimodal conditions. In Exp. 4, we addressed this possibility by assessing the worse uni condition in one session and then comparing performance in that condition to performance on the cross condition in subsequent, independent sessions. Third, the task in this experiment was changed to a "more/fewer" judgment in order to avoid the potential complications of the "same/different" task. Similar results in this experiment, despite the use of a different task, would provide some evidence for the generality of the comparison abilities we observed in Exps. 1, 2, and 3.

Method

Participants. One male and ten females between the ages of 18 and 35 participated in the study. All had normal or corrected-to-normal hearing and vision. Three additional subjects were excluded for failing to complete all 3 experimental sessions.

Design. All participants completed 3 separate experimental sessions, with conditions presented in counterbalanced order. Auditory/Temporal, Visual/Spatial, and Crossmodal/Cross-format (again abbreviated to "Cross") conditions were blocked; participants received two blocks of 40 trials of each condition, for a total of 80 trials per condition per session (and a grand total of 240 trials per condition). Before the experimental blocks began, there were 10 practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block. There were 5 possible comparison ratios, each presented in 4 absolute set sizes as shown in Table 1. In each block, each ratio/set size combination was presented twice (once in the order in which is shown in the table, and once in the reverse order).

| | Set Size 1 | Set Size 2 | Set Size 3 | Set Size 4 |
|---|---|---|---|---|
| **Ratio 1:2** | 10 and 20 | 15 and 30 | 20 and 40 | 25 and 50 |
| **Ratio 2:3** | 10 and 15 | 20 and 30 | 24 and 36 | 30 and 45 |
| **Ratio 3:4** | 9 and 12 | 15 and 20 | 24 and 32 | 30 and 40 |
| **Ratio 4:5** | 12 and 15 | 20 and 25 | 28 and 35 | 40 and 50 |
| **Ratio 7:8** | 14 and 16 | 21 and 24 | 28 and 32 | 35 and 40 |

Table 1. The numerosities used in Experiment 4.

Stimuli. Experiment 4 used the same basic stimuli as Experiment 3, visual arrays and auditory sequences, except that their numerosities were determined as shown above.

Procedure. The procedure was the same as that used in Experiment 3, except that participants were now required to judge the numerosity of the second stimulus as "more" or "fewer" than the first. Auditory sequences were also altered so that tone durations varied from 20 – 60 ms within each sequence.

Measures. The method we used previously to assess Cross performance compared to Worse Uni may have introduced a bias by yielding a deceptively low Worse Uni score. This was due to the fact that we used the same data set both to determine which "Uni" condition the subject was worse at, Auditory/Temporal or Visual/Spatial, and to provide the accuracy score used in the comparison with Cross accuracy as well. Because each participant in Experiment 4 completed 3 sessions, we were able to avoid this bias in the present analysis by using an independent data set for each subject to determine which Uni condition led to lower performance for that subject. We chose subjects' Worse Uni conditions from data from that subject's first session, but used accuracy scores only from the last 2 sessions for the comparison.

Results.

The mean Worse Uni accuracy score was 81%, and the mean Cross accuracy score was 77% (see Figure 5) This slight accuracy difference between Cross and Worse Uni trials did not reach significance ($t(10) = -2.11$, $p > .05$). A repeated-measures ANOVA of correct RTs (Condition x Ratio x Group Size) revealed main effects of all three within-subjects factors [Condition: $F(2, 18) = 17.3$, $p < .0005$; Ratio: $F(4, 36) = 7.8$, $p < .0005$; Size: $F(3, 27) = 3.8$, $p < .05$]. The main effect of Size results from a *decreasing* linear trend in reaction time as set size increases, while reaction time *increases* with Ratio. Regressions are shown in Figure 6 for Size and in Figure 7 for Ratio.

Discussion.

The key result of Experiment 4 is that response time does not increase at all with absolute set size; comparison ratio alone determines the time necessary for these numerosity judgments. This finding is inconsistent with theories of numerosity estimation that rely upon iterative mechanisms, because such mechanisms would necessarily require more time to enumerate larger sets. If anything, participants in the present experiment were faster for larger set sizes. We must be sure to take into account, however, the complications that temporal sequence trials introduce to any measure of reaction time, as discussed in the earlier experiments. Therefore, strong conclusions about enumeration mechanisms should not be drawn based solely on reaction time data from trials involving sequential stimuli. However, Exp. 4 also provides RT data from trials which involve *only* visual arrays, which do not introduce the complications that sequences do. Any claim to be made regarding mechanisms of enumeration should rest upon RT data from these Visual trials. The Visual trials show no increase in RT with increasing set size, so this experiment can provide strong evidence of a non-iterative mechanism for the enumeration of spatial/simultaneous sets. However, temporal/sequential sets may be enumerated through a different process.

This experiment also supported our previous findings that adults can very effectively compare numerosities across both modality and format. In Experiment 4, participants again showed little or no performance cost when comparing numerosities in the Cross condition relative to performance on their worse single modality/format condition, even though the Cross comparisons were again being made across both modality and format. There may be some slight improvement of Cross performance relative to worse single modality/format in Exp. 4 (compared to Exp. 3), for several reasons. First, Experiment 4 included many more trials (240 per condition for each subject), so participants had more practice in Exp. 4. It is possible that this additional practice produced more of an improvement in the Cross condition than it did in the single modality/format conditions. Second, the analysis we used in order to correct for our biased analysis in Experiment 3 may have affected the result in an unexpected way. Though the newer method was expected, if anything, to increase the difference between crossmodal and worse unimodal performance, perhaps it had the opposite effect. This would be possible if practice effects for this task had a greater influence on the condition that participants were worse at to begin with[4]. A third possible explanation is the fact that the task was changed from "choose same or different" in Exp. 3 to "choose fewer or more" in Exp. 4. The same/different task is much less clear-cut, and the difficulties involved in making this judgment might be especially pronounced for the Crossmodal condition.

---

[4] For example, if a participant performed badly in the "visual" condition in the first session, and moderately in the "auditory" condition, then practice might improve "visual" performance a great deal, but "auditory" performance only slightly. Yet "visual" would be counted as the "worse" unimodal score. However, this does not seem to be the cause of the difference between Experiments 3 and 4, because participants were very consistent across sessions in that a person who performed better with visual arrays in the first session tended to continue that way throughout the study.

There is another possible explanation which does not involve the actual comparison component of the task. All conditions were blocked in Experiments 1, 2, and 3, so that the Crossmodal and Cross-format conditions were the only ones in which different kinds of stimuli were presented. Consider the task-switching involved in the "Crossmodal" condition of Experiment 3: at the beginning of each trial, the participant knows neither the format nor the modality of the next stimulus. It could be a sequence or an array; auditory or visual. In Experiment 1, however, modality is unknown but everything is a sequence, and in Experiment 2, format is unknown but everything is visual. It is possible that preparing for each trial is more difficult when both of these quantities are unknown. To test this possibility, we repeated a briefer version of Experiment 4 (40 trials per condition, rather than the 240 trials per condition of Experiment 4), with the Auditory/Temporal and Visual/Spatial trials interleaved. The task was again a more/fewer decision, so that the difficulties of the same/different task could not affect performance. Results showed that the Crossmodal/Crossformat performance deficit remained, despite this interleaved design, suggesting that task-switching problems did not contribute to the slight cost for the Crossmodal/Crossformat condition. We conducted a number of follow-up studies using the more/fewer task with various trial and block structures, and the Crossmodal/Crossformat condition always produced a slight deficit. This suggests that the slight deficit we observed in Exp.4 is a robust result, and that it is not due to the same/different task of Exp. 3.

Experiment 4's results support our findings in Exp. 3 that crossmodal/cross-format comparisons are only slightly more difficult than intramodal or intra-format comparisons. Again, there was a non-significant drop in accuracy when comparing across both format and modality simultaneously. The experiment also shows that because response time does not increase with set size for the Visual condition, the mechanism used to enumerate spatial arrays is likely to be non-iterative.

## 2.5 General Discussion

The present finding that there is little or no cost for comparing numerosities across stimulus format or modality, relative to accuracy on intramodal and intra-format comparisons, shows that adults' judgments of approximate numerosity are based on abstract representations of number. In addition, adults take no longer to make comparisons between large visual sets than between small visual sets, when the ratio between the numerosities to be compared remains the same. Human adults appear to compare large discrete spatial quantities through the non-iterative construction of representations of numerosity that are independent of the modality or format of the stimulus.

Our findings show that these numerosity judgments could not have been made on the basis of modality- and format-specific stimulus attributes such as duration, rate, texture density, or area. Rather, these quantities may have acted as cues in the formation of an abstract numerosity representation. Durgin (1995) has shown that some models of human numerosity perception which purport to depend only on one stimulus attribute, such as area, must in fact implicitly make use of density information if they are to deal with a range of numerosities. Related studies have suggested that numerosity judgments are influenced by different stimulus attributes depending on the range of numerosities to be judged (Durgin, 1995). The present experiments show that representations of perceptual stimulus attributes cannot be directly responsible for numerosity judgments, and that there must be some transformation of this perceptual information into an abstract form. Taken together, these findings provide strong evidence that abstract numerosity representations are constructed from multiple perceptual cues, much as a unified percept of depth is the product of many cues such as texture gradients, binocular disparity, and motion parallax.

Our second major finding is that abstract numerosity representations appear to be derived from perceptual representations by a non-iterative process, at least when the quantity to be enumerated is presented in a spatial/simultaneous format. How can we reconcile the current results with other findings that have been presented in support of iterative enumeration mechanisms? A long tradition of evidence from numerosity estimation tasks has led some researchers to posit iterative enumeration mechanisms such as preverbal counting(Gallistel and Gelman 1992), or protocounting (Davis and Perusse 1988). When adults are shown an array of dots, for example, and asked to make a speeded judgment of how many there are (by producing a number), response time reliably increases with the number of dots; this has been explained in terms of serial enumeration mechanisms. However, as the number of dots increases, our rough approximation of their numerosity becomes even rougher, and our fuzzy representation of the number of dots maps onto a larger set of possible symbolic responses. This increase in the number of response options could make response selection more difficult, accounting for the observed increases in reaction time. In the task used in Experiment 4, on the other hand, responses were limited to two alternatives, and we observed no increase in reaction time as a function of set size. A preverbal counting system could not have produced the patterns of reaction time we observed in Experiment 4; the fact that reaction time did not increase with set size shows that these sets were enumerated by a non-iterative process. Support for non-iterative enumeration processes in children has been found as well (Huntley-Fenner 2001).

Our claim that a non-iterative process is involved in deriving numerosity meshes well with prior findings that numerosity judgments are extremely sensitive to perceptual properties of the stimuli to be enumerated. It is certainly possible for an iterative counting-like process to be affected by such properties. For example, any counting mechanism must individuate elements, and the effectiveness of this individuation process could certainly be influenced by changes in the arrangement or size of the elements. But the stimulus-dependence that has been observed can be more parsimoniously attributed to the operation of a non-iterative enumeration process, which uses multiple stimulus attributes in combination as cues to numerosity. Thus, information about a visual array such as its density and area might be transformed into a representation of the array's numerosity that is modality-independent. Under such conditions, any attribute of the stimulus that affects our perception of these cues (e.g. density aftereffects or anchoring effects) will clearly alter our representations of numerosity as well. The nature of this transformation from stimulus-specific properties to numerosity representations remains unknown; the same mechanism may serve to convert all spatially presented stimuli to abstract form, while another may perform the same task for all temporal presentations. The latter sort of mechanism may be responsible for animals' representation of the numerosity of a sequence of events; it has been proposed that animals keep track of the average interval between events and the overall sequence duration, using these two durations to compute the total number of events (Church and Broadbent 1990).

The task involved in Experiment 4 required a more vs. fewer judgment, which is of course a judgment of *relative* numerosity. Some investigators have suggested that absolute and relative numerosity judgments should be considered separately, and that animals' ability to make relative judgments has nothing to do with the purely human ability to judge absolute numerosity (Davis 1993). Therefore it could be argued that our results are not truly comparable to those that require estimation of a single quantity, for example, which are precisely those results that lend support to the idea of an iterative nonverbal counting procedure (Gallistel and Gelman 1992)

However, in order to discuss the presence or absence of various forms of numerical competence, it is necessary to define exactly what is meant by "competence with absolute numerosity." A nonhuman animal may have no need of absolute numerosity judgment for foraging, where the goal is to identify which food source yields "more" than the others. Yet in order to judge relative numerosity, there must be some implicit representation of absolute numerosity first unless animals directly extract differences or ratios. Absolute numerosity judgment may also be useful in judging whether a particular foraging effort is worthwhile in the first place. One possible definition of 'numerical competence' includes the implicit requirement that the quantity in question must be accessible to the animal and available as a modulator of behavior. It is possible that nonhuman animals do not have access to any representation of absolute numerosity; therefore, according to the above definition, they would not demonstrate numerical competence for judgments of absolute numerosity. If, on the other hand, we speak of nonhuman animals' ability to *represent* absolute numerosity, making no claim about the availability of this representation as a guide for behavior, then it is clear that this ability is present. Indeed, the fact that chimpanzees can be taught to use absolute number at all (with extensive training: Matsuzawa 1985; Kawai and Matsuzawa 2000) suggests that they may have the ability to represent it all along, though it is also possible that extensive training gives them previously unavailable representational power.

Future work will test whether our findings with relative numerosity judgments can be found using tasks that necessitate the explicit use of absolute approximate numerosities, without requiring manipulation of Arabic digits or choice among large numbers of responses. In addition, though we have provided evidence in support of a general class of enumeration mechanisms, we have not identified the specific steps involved. Is there a unitary mechanism responsible for the enumeration of all elements, regardless of modality or format? Or is the process for the most part modality- and format-specific, though it culminates in a representation that is neither? Ongoing studies will target the nature of the process by which modality-specific stimulus properties serve as cues in the non-iterative formation of abstract representations of numerosity.

## Acknowledgments

Hilary Barth, Nancy Kanwisher, and Elizabeth Spelke, Department of Brain and Cognitive Sciences at the Massachusetts Institute of Technology.

This research was supported by National Institute of Health grant MH56037 to N. Kanwisher, National Institute of Health grant R37 HD23103 to E. Spelke, and National Science Foundation ROLE grant REC-0087721 to N. Kanwisher and E. Spelke.

A version of this chapter is under review for *Cognition.*

References


Allik, J. and T. Tuulmets (1991). "Occupancy model of perceived numerosity." Perception & Psychophysics **49**(4): 303-314.

Allik, J. and T. Tuulmets (1993). "Perceived numerosity of spatiotemporal events." Perception & Psychophysics **53**(4): 450-9.

Allik, J., T. Tuulmets, and P. G. Vos (1991). "Size invariance in visual number discrimination." Psychological Research **53**(4): 290-295.

Church, R. M. and H. A. Broadbent (1990). "Alternative representations of time, number, and rate." Cognition **37**(1-2): 55-81.

Church, R. M. and W. H. Meck (1984). The numerical attribute of stimuli. Animal Cognition. H. L. Roitblatt, T. G. Bever and H. S. Terrace. Hillsdale, NJ, Erlbaum: 445-464.

Davis, H. (1993). Numerical competence in animals: Life beyond Clever Hans. The development of numerical competence: Animal and human models. Comparative cognition and neuroscience. E. J. C. Sarah Till Boysen, Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US: 109-125.

Davis, H. and R. Perusse (1988). "Numerical competence in animals: Definitional issues, current evidence, and a new research agenda." Behavioral and Brain Sciences **11**(4): 561-615.

Dehaene, S., G. Dehaene-Lambertz, and L. Cohen. (1998). "Abstract representations of numbers in the animal and human brain." Trends Neurosci **21**(8): 355-61.

Durgin, F. H. (1995). "Texture density adaptation and the perceived numerosity and distribution of texture." Journal of Experimental Psychology: Human Perception & Performance **21**(1): 149-169.

Gallistel, C. R. and R. Gelman (1992). "Preverbal and verbal counting and computation. Special Issue: Numerical cognition." Cognition **44**(1-2): 43-74.

Ginsburg, N. (1991). "Numerosity estimation as a function of stimulus organization." Perception **20**(5): 681-686.

Ginsburg, N. and A. Nicholls (1988). "Perceived numerosity as a function of item size." Perceptual & Motor Skills **67**(2): 656-658.

Ginsburg, N. and L. Pringle (1988). "Haptic numerosity perception: Effect of item arrangement." American Journal of Psychology **101**(1): 131-133.

Huntley-Fenner, G. (2001). "Children's understanding of number is similar to adults' and rats': numerical estimation by 5-7-year-olds." Cognition **78**(3): B27-40.

Lechelt, E. C. (1975). "Temporal numerosity discrimination: intermodal comparisons revisited." Br J Psychol **66**(1): 101-8.

Massaro, D. W. (1976). "Perceiving and counting sounds." J Exp Psychol [Hum Percept] **2**(3): 337-46.

Mix, K. S., J. Huttenlocher, and S. C. Levine(1996). "Do preschool children recognize auditory-visual numerical correspondences?" Child Dev **67**(4): 1592-608.

Mix, K. S., S. C. Levine, and J. Huttenlocher (1997). "Numerical abstraction in infants: another look." Dev Psychol **33**(3): 423-8.

Moore, D., J. Benenson, J. S. Reznick, and M. Peterson (1987). "Effect of auditory numerical information on infants' looking behavior: Contradictory evidence." Developmental Psychology **23**(5): 665-670.

Moyer, R. S. and T. K. Landauer (1967). "Time required for judgements of numerical inequality." Nature **215**(109): 1519-20.

Shams, L., Y. Kamitani, and S. Shimojo (2000). "Illusions. What you see is what you hear." Nature **408**(6814): 788.

Starkey, P., E. S. Spelke, and R. Gelman (1983). "Detection of intermodal numerical correspondences by human infants." Science **222**(4620): 179-181.

Starkey, P., E. S. Spelke, and R. Gelman (1990). "Numerical abstraction by human infants." Cognition **36**(2): 97-127.

**Figure 1a**. Schematic depiction of possible ways to perform a numerical comparison task within modality – such tasks cannot rule out the possibility that modality-specific representations are being compared. The question mark is an unknown perceptual representation of numerosity, which is specific to some property of the stimulus such as modality. The graduated cylinder is a numerical representation that is not modality-specific.

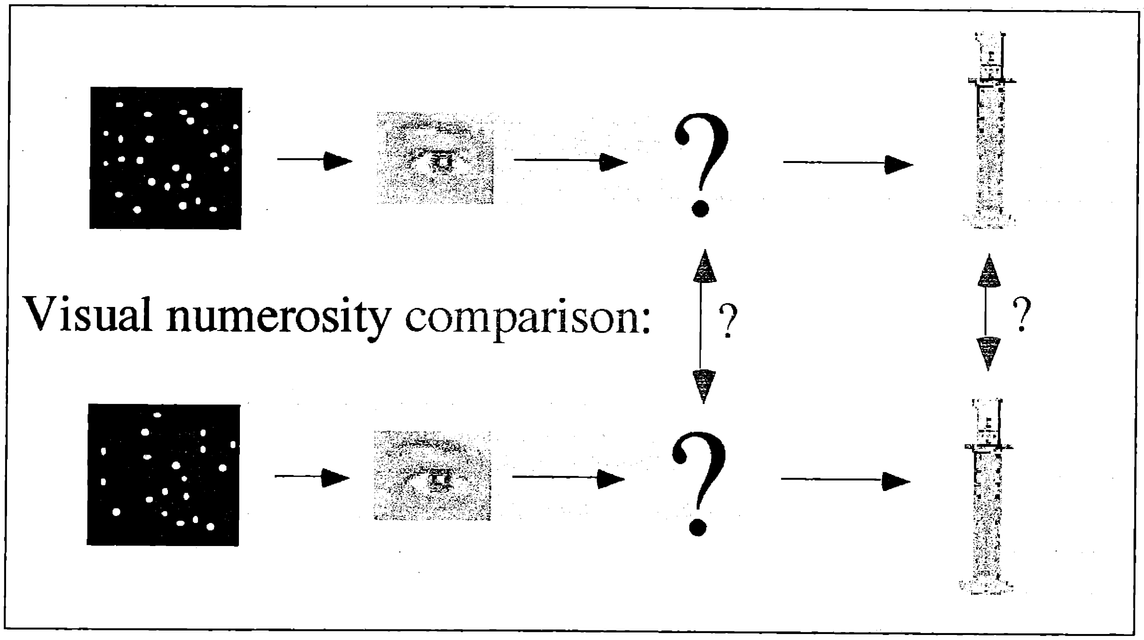



**Figure 1b.** Crossmodal numerical comparisons cannot be made on the basis of modality-specific representations.

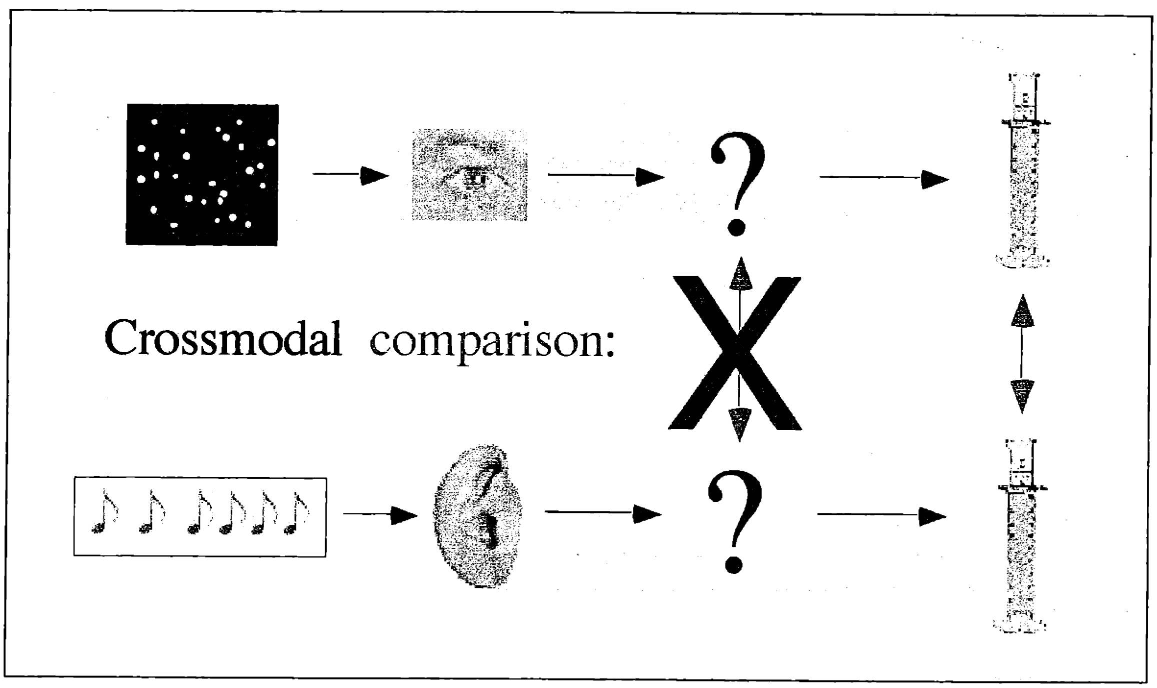


**Figure 2.** Accuracy scores for Experiment 2.1

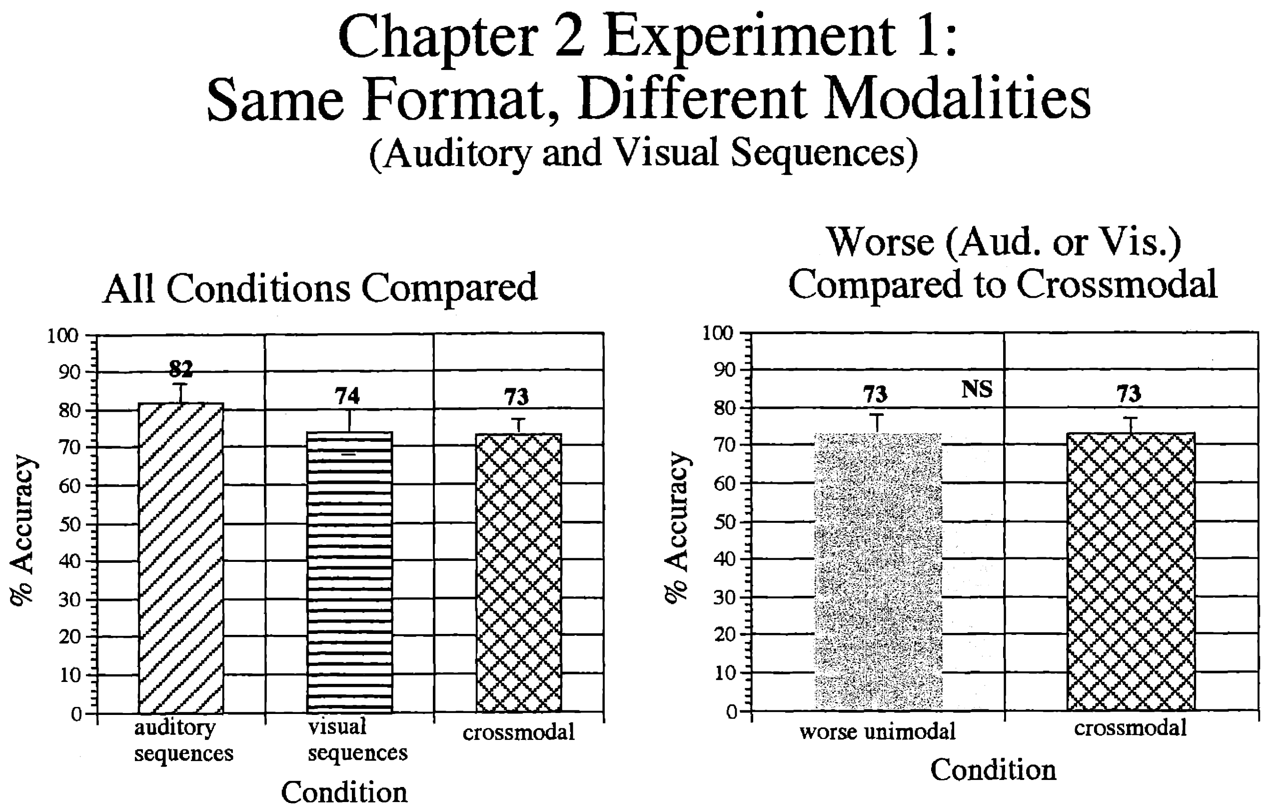


**Figure 2a.** Accuracy scores as function of ratio for Experiment 2.1

# Chapter 2 Experiment 1: Accuracy: Different Comparison Ratios

(Auditory and Visual Sequences)

Worse (Aud. or Vis.) Compared to Crossmodal
"Different" trials only

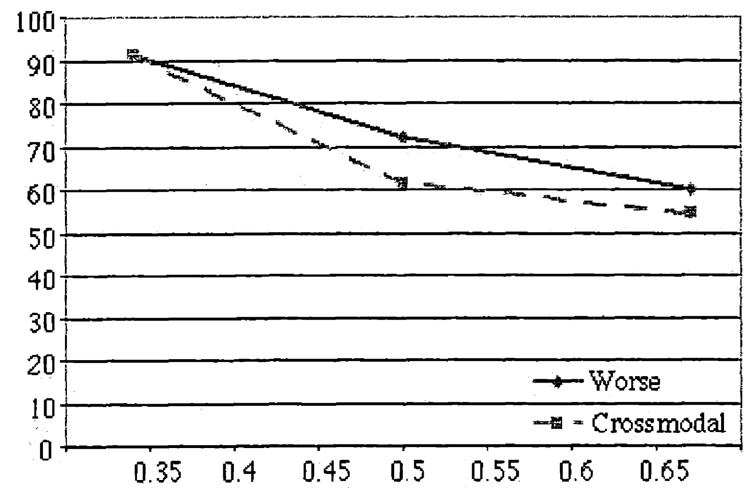


**Figure 3.** Accuracy scores for Experiment 2.2

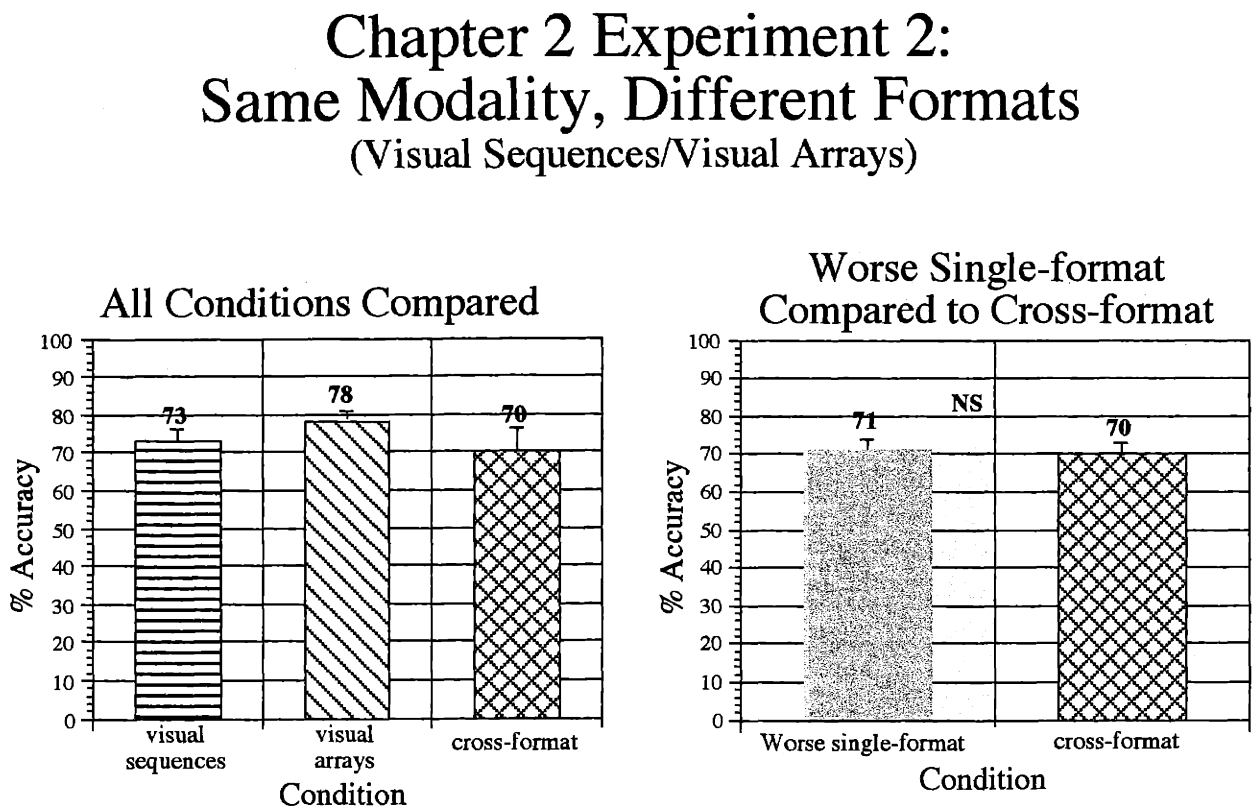


**Figure 3a.** Accuracy scores as function of ratio for Experiment 2.2

# Chapter 2 Experiment 2: Accuracy: Different Comparison Ratios

(Visual Sequences/Visual Arrays)

Worse (Sequence or Array) Compared to Crossmodal
"Different" trials only

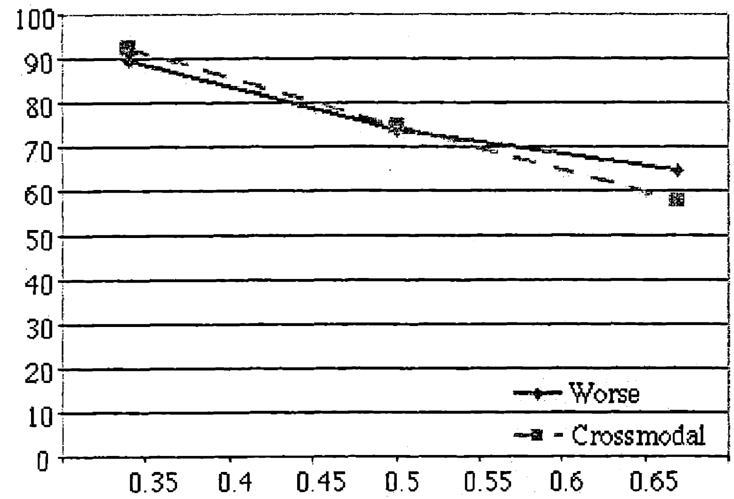


**Figure 4.** Accuracy scores for Experiment 2.3

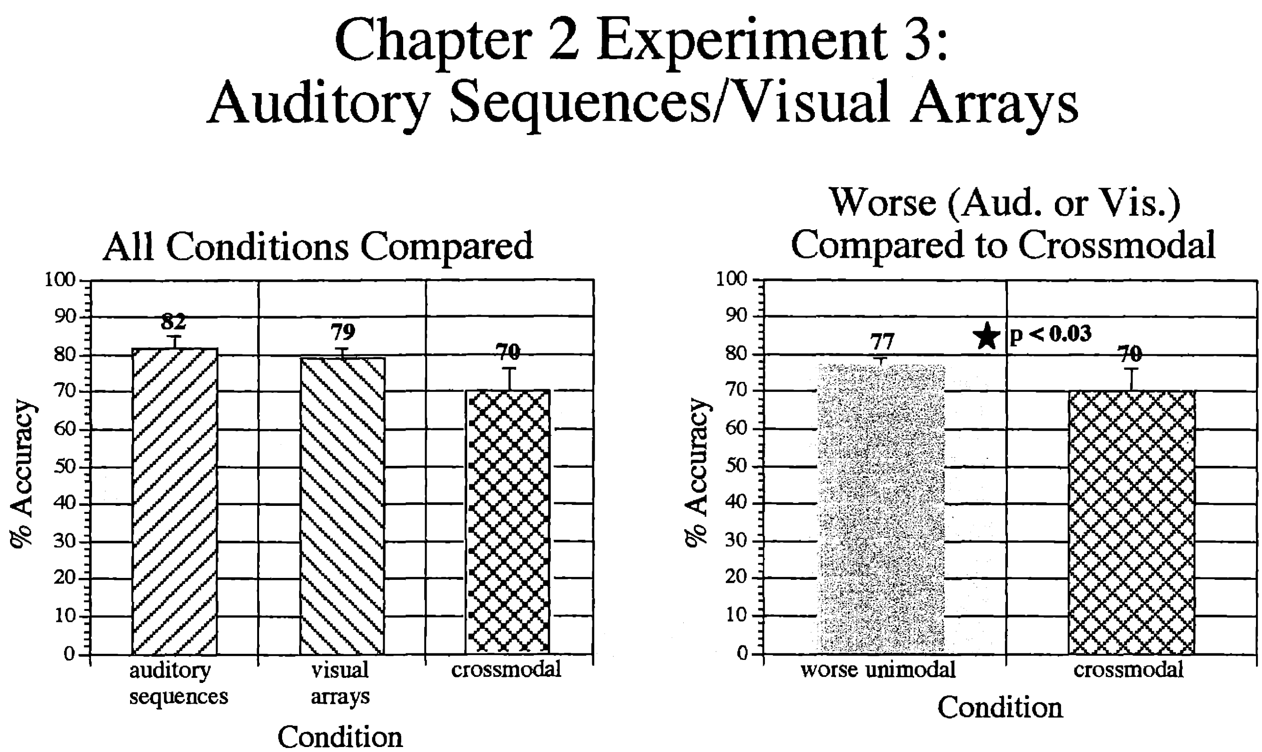


**Figure 4a.** Accuracy scores as function of ratio for Experiment 2.3

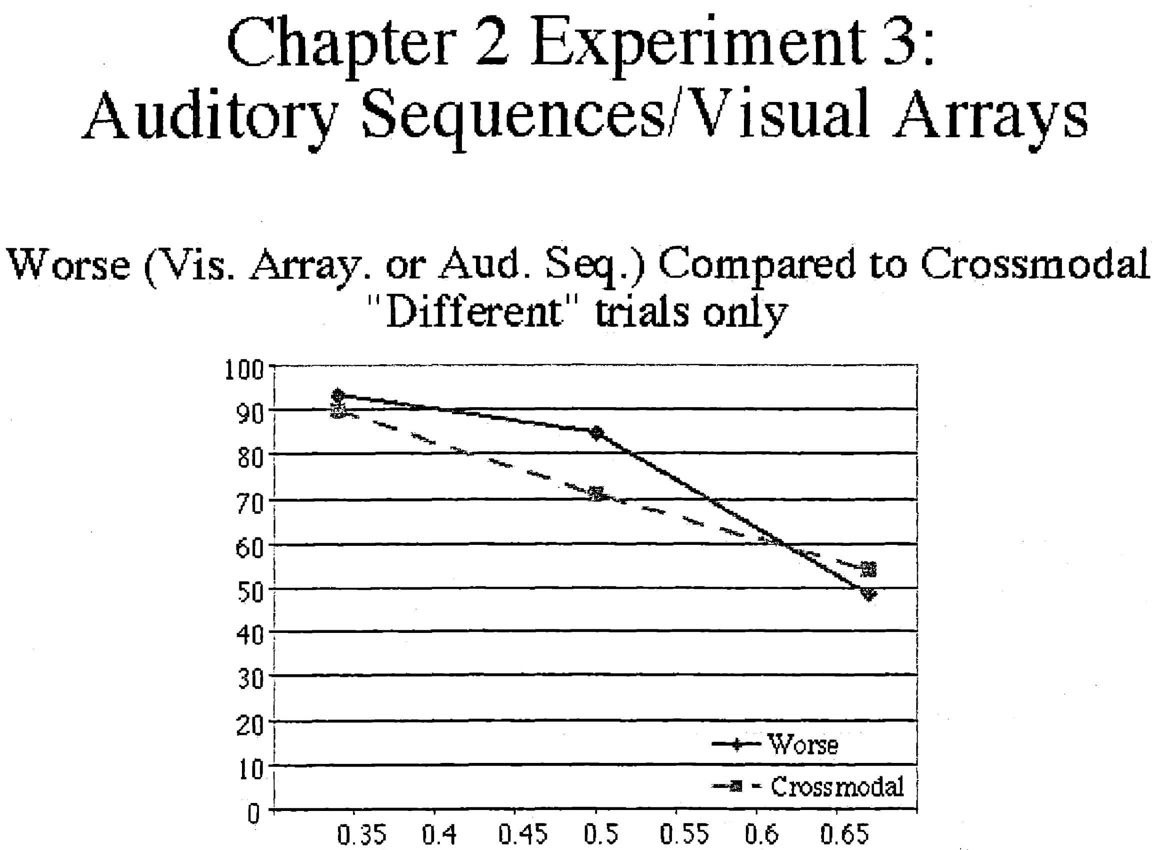


**Figure 5.** Accuracy scores for Experiment 2.4

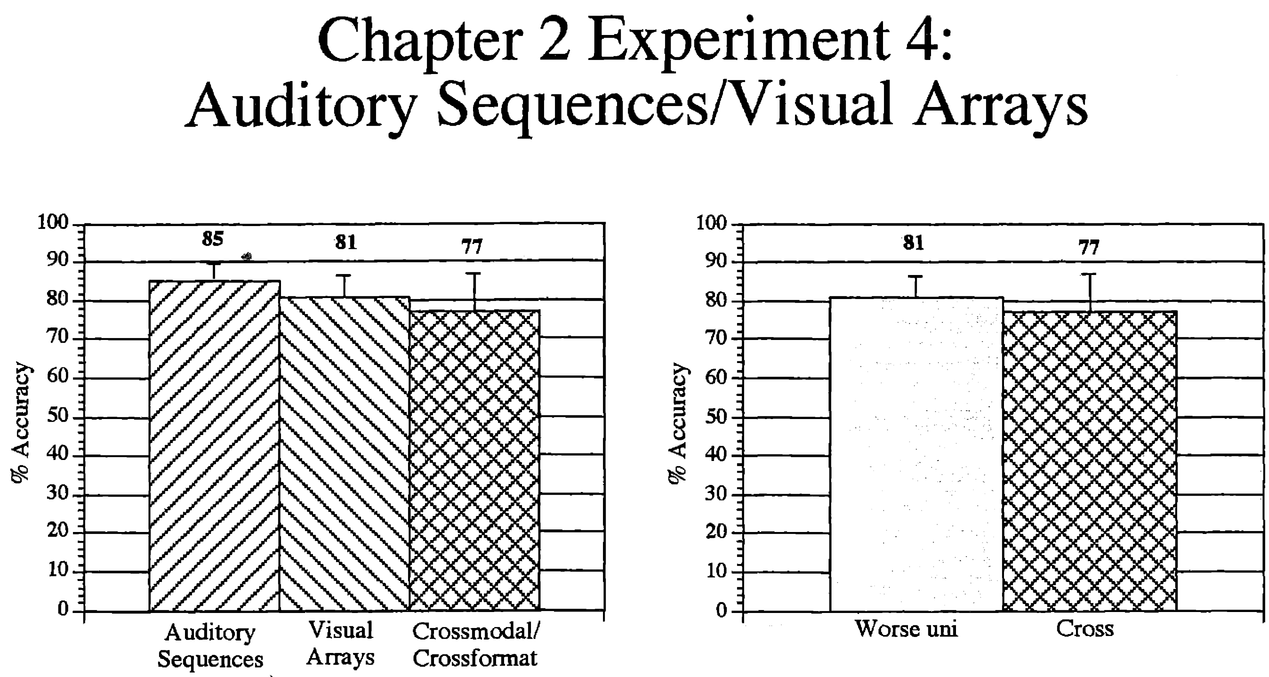
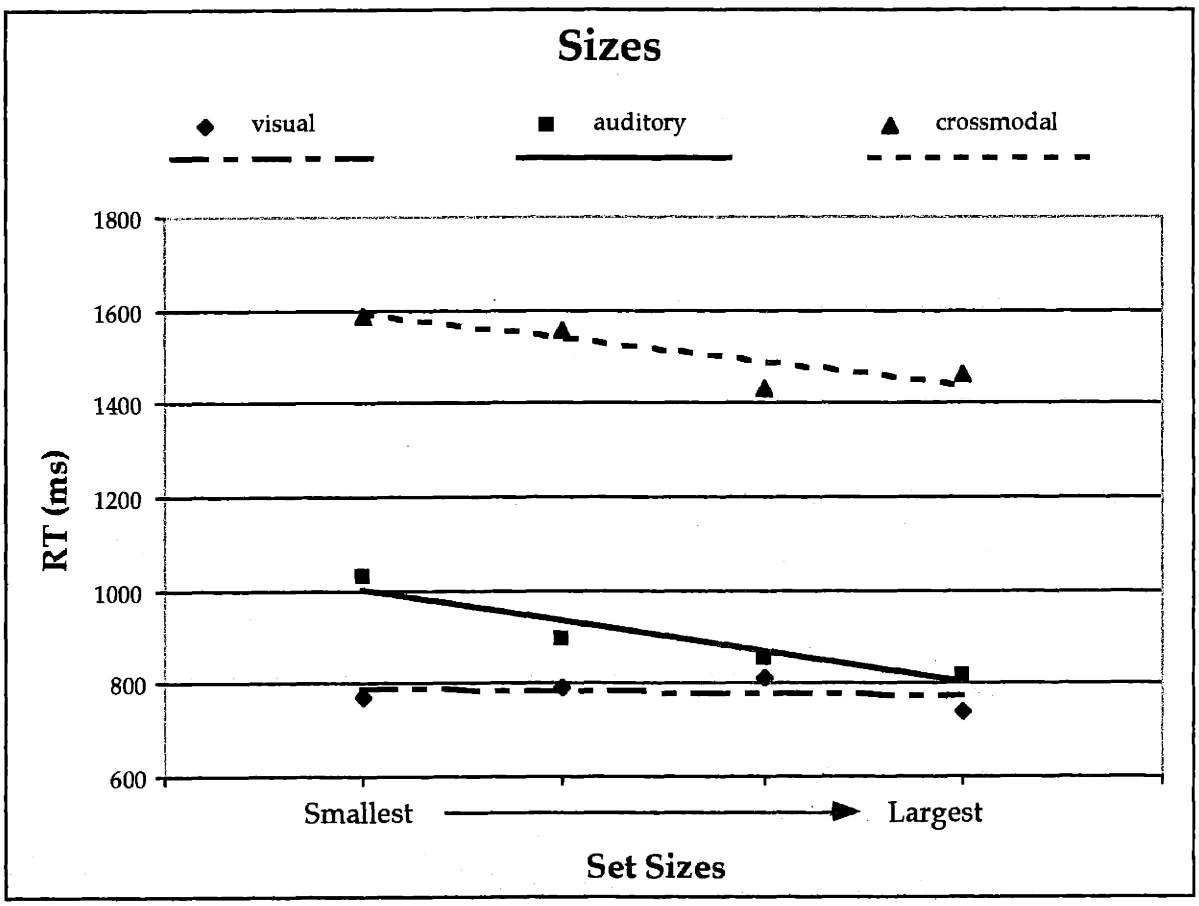


**Figure 6.** Regression lines for RT as a function of set size in Experiment 2.4



# Chapter 3

## Nonsymbolic Arithmetic with Large Approximate Numerosities


Abstract

Human adults, like infants and nonhuman animals, are able to process numerical quantities without the use of symbolic notation; this ability has been called the "number sense." Previous research has elucidated a great deal about the abstract nature of this number concept and the form of the representations from which it springs. One question which remains unanswered concerns the depth and richness of the number sense in humans and animals: can number representations function in arithmetic operations? Some evidence has suggested that the number sense constitutes a biologically specified knowledge system, shared by humans and many other species, which underlies the uniquely human mathematical capabilities we have developed. Results suggesting that nonhuman animals are capable of some arithmetic computations are consistent with this view, but little is known about comparable abilities in humans. The present experiments assess the abilities of human adults to perform computations on large nonsymbolic sets. We find that large approximate quantity representations are indeed available for operations such as addition and subtraction, multiplication and division. We discuss the implications of the specific patterns of results for the relationship between symbolic and nonsymbolic arithmetic, and for the nature of large number representation.

## 3.0 Introduction

In the past few decades, researchers have demonstrated that human infants and nonhuman animals have impressive numerical abilities that share many characteristics with those observed in human adults. This has led to the common supposition that there is a basic preverbal "number sense," or rough representation of mental magnitude, that is present across species. These mental magnitudes resemble the representations we use for other physical magnitudes, such as weight or brightness, in that they are subject to Weber's Law (the discriminability of two magnitudes depends on their ratio). Evidence from studies on animals and humans further suggests that this mental magnitude representation may form the basis for more complex symbolic numerical capabilities developed by humans alone (Dehaene, Dehaene-Lambertz and Cohen 1998). However, it remains to be seen whether approximate numerical representations can be used in arithmetical manipulations themselves. If it is true that the preverbal number sense is the biologically specified base of the symbolic numerical abilities developed by humans, we would expect preverbal representations to be available for arithmetical manipulation.

There is certainly evidence that the preverbal number sense underlies, or at the very least accompanies, some kinds of uniquely human numerical abilities. Many tasks that deal explicitly with exact numbers may actually make implicit use of approximate quantity representations, which appear to be automatically activated in most tasks that involve numbers even when numerical value is irrelevant. For example, when human adults are asked to compare the numerical sizes of two Arabic digits, their speed and accuracy depend on the ratio of the two numbers. This is a characteristic result for operations that are subject to Weber's Law.

Approximate quantity manipulations may be involved in more aspects of simple arithmetic than we realize. When children are first learning to subtract, they often use a strategy referred to as the "choice" algorithm, which involves either counting from the subtrahend up to the minuend (e.g. to answer "8 - 5" by counting "6, 7, 8" which yields the answer in the number of counting-steps, 3) or counting from the minuend down by a number of steps equal to the subtrahend (e.g. to answer "8 - 3" by counting "7, 6, 5" which yields answer in the final number reached, 5). The "choice" algorithm is so named because children choose the strategy that requires the smallest number of counting-steps. That is, before they can make their choice of the most efficient counting method, they must first compute the answer, very possibly through nonverbal magnitude representations (Resnick 1983; Gallistel and Gelman 1992). Evidence for the intertwining of approximate numerical ability with symbolic math is not limited to the early stages of mathematical training, however; adults are able to perform rapid approximate arithmetic on Arabic digits, apparently utilizing a cerebral circuit distinct from that involved with the retrieval of exact numerical facts (Dehaene, Spelke, Pinel, Stanescu, and Tsivkin 1999). However, we do not know what kinds of operations adults can perform on quantities that are only approximately specified in the first place. If an innately determined number sense, common to many species, does in fact form the basis for more complex human mathematical skills, then approximate operations available to animals and preverbal humans could be expected to appear in human adults as well. Also, if the preverbal numerical abilities that we share with other animals underlie later, more sophisticated operations like symbolic arithmetic, we might expect to see patterns of performance in nonsymbolic arithmetic tasks that reflect the patterns observed in symbolic tasks.

Human infants can perform simple arithmetic operations such as addition and subtraction at least on small numbers of objects. When presented with displays in which 2 items are placed behind a screen and then one is removed, infants expect exactly one item to remain. Similarly, when two items are sequentially placed behind a screen, infants expect there to be exactly two (Wynn 1992). Studies with various nonhuman animals also demonstrate remarkable arithmetic skills though the subjects cannot make use of a symbolic number system (Church and Meck 1984). Though most of these studies involve extensive training, wild rhesus monkeys have been found capable of the same simple addition and subtraction operations that Wynn demonstrated in human infants, and more (Hauser, MacNeilage, and Ware 1996; Sulkowski and Hauser 2001). While alternative perceptual explanations have not been entirely ruled out, it seems that there is promising evidence for arithmetic capabilities in some primates.

These particular infant and monkey studies involved the manipulation of exact very small numerical quantities. As there is some evidence that processing of such quantities may be distinct from that of larger approximate numerosities (Dehaene 1997; Carey 2001; but see Gallistel 1990), these studies may not deal with mechanisms analogous to those that could allow human adults to perform approximate arithmetical operations. While infants can discriminate between larger sets as well, given large enough comparison ratios (Xu and Spelke 2000), it is not clear whether these ranges of numerosities may also be used in arithmetic operations. However, there is extensive evidence that animals process representations of larger numbers arithmetically. Research on foraging behavior has shown that foraging patterns seem to depend on animals' calculation of rate of return, which requires a complex combination of duration and number information (Gallistel and Gelman 1992). A large body of research on temporal processing, which is likely to depend upon representations very similar to representations of numerical quantity, has demonstrated arithmetic processing of durations (Meck, Church, and Gibbon 1985; Gibbon, Malapani, Dale, and Gallistel 1997).

More direct assessments of arithmetic capabilities with larger quantities have also produced convincing results, many of which suggest that certain arithmetic operations such as summation may be automatic when two quantities are presented closely in space and/or time. In a striking demonstration of spontaneous summation across modalities, researchers showed that when rats see two lights and hear two sounds at once, they give the conditioned response that means "4" (Church and Meck 1984). Spontaneous summation has also been demonstrated in many other species; chimpanzees have been found to sum quantities presented both nonsymbolically (without training) and symbolically (with training of the symbolic associations, but no training of the summation process) (Rumbaugh, Savage-Rumbaugh, and Hegel 1987; Rumbaugh, Savage-Rumbaugh, and Pate 1988; Boysen and Berntson 1989). Chimpanzees are not the only animals capable of symbolic operations like addition; spontaneous summation of trained symbols has also been demonstrated by Olthof and colleagues in squirrel monkeys and in pigeons (Olthof, Iden, and Roberts 1997; Olthof and Roberts 2000).

Characteristics of arithmetic processing may be quite stable even across such seemingly different operations as learned symbolic addition in human adults and rough addition of quantities in animals. By the time formally educated humans reach adulthood, they tend to perform simple addition and multiplication tasks by retrieving learned facts (Ashcraft 1992). They also tend to sum small symbolic quantities automatically, as it seems nonhuman animals do with nonsymbolic quantities (LeFevre, Bisanz, and Mrkonjic 1988; Stadler, Geary, and Hogan 2001). The "obligatory activation" of the sum of two Arabic numerals has been explained in terms of the overlearned associations between number pairs and their sums in adult humans with years of training in simple arithmetic. Given the similarity of automatic summation in animals, though, it is tempting to suppose that this symbolic-learning explanation is mistaken, and that a common process underlies both phenomena.

In this series of studies, we probe adults' performance on the addition, subtraction, multiplication, and division of large nonsymbolic quantities. It remains to be seen whether nonsymbolic large-number arithmetic in human adults will mirror the performance patterns observed in other species, other developmental stages, or other types of tasks.

## 3.1 Chapter 3 Experiment 1

In Experiment 1, we asked participants to perform comparison and addition tasks on large sets. Comparisons could be made within a type of set (comparing the numerosities of 2 dot arrays) or across types of sets (comparing a dot array to a tone sequence). Similarly, addition could require adding 2 dot arrays or adding a dot array to a tone sequence. One might expect performance to be worse in the Addition conditions for many reasons. First, in the Addition tasks, participants had to perform approximate additions of large sets and then keep track of the mental representation of this sum, in order to compare it to the test set. The Addition conditions require a comparison between the sum (a representation that is the result of an operation on two explicitly presented sets) and the test set (which is also explicitly presented). In the Comparison tasks, participants only need to compare two explicitly presented sets. Also, representations of approximate numerosity are noisy; the inclusion of a third set should mean that the Addition tasks would allow less accurate judgments than the Comparison tasks, which only require the subject to deal with two sets. And finally, participants might simply be unable to perform an operation like addition on large sets of elements, within or across modalities.

*Method*

Participants. Fourteen subjects between the ages of 18 and 35 were paid to participate in the study. All had normal or corrected-to-normal hearing and vision.

Apparatus. Participants sat in a small darkened room at a distance of approximately 60 cm from the presentation screen. Visual stimuli were presented on a Sony Multiscan monitor by a Power Macintosh 8600 computer. Auditory stimuli were presented from the Macintosh's built-in speaker. The apparatus was the same for all of the experiments that follow.

Design. All participants received the same four stimulus conditions in counterbalanced order. Visual Comparison, Visual Addition, Crossmodal Comparison, and Crossmodal Addition trials were blocked; participants received one block of 62 trials of each condition. Before the experimental blocks began, there were 5 practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block (Visual Comparison, Visual Addition, Crossmodal Comparison, or Crossmodal Addition). The numerosities used in the Comparison task were matched to the comparisons to be made in the Addition task. For example, an addition problem might be an approximate version of "30 + 20 = (50) vs. 40." The corresponding comparison problem would be "50 vs. 40." There were four possible ratios of the numerosities to be compared: 0.75, 0.8, 0.83, and 0.86.

Stimuli. An auditory sequence consisted of a series of brief tones. All tones in a particular sequence had the same duration, but tone duration varied from sequence to sequence between ~20 and ~60 ms. Tone presentation rate ranged from 7 to 11 tones per second, varying randomly so that the duration of the entire sequence was not a reliable cue to numerosity. The end of each tone sequence was marked by a brief high-pitched beep. A visual array consisted of small black dots on a mid-gray background. The arrays of dots were presented inside an imaginary square measuring ~13 by 13 cm. The distribution of the dots was pseudorandom, though they did not touch or overlap. All of the dots in a particular array were the same size, but dot diameter varied from array to array between 0.2 and 0.6 cm.

Procedure. In the Visual Comparison condition, a small red fixation cross appeared for ~400 ms, followed by a dot array, a pause of ~930 ms, and a second dot array. The task was to determine whether the second array had more or fewer dots. In the Crossmodal Comparison condition, one dot array was replaced by a tone sequence; participants determined whether the second set (whether it was a dot array or a tone sequence) contained more or fewer dots. The Addition conditions were somewhat more complex; in the Visual Addition case, participants first saw a small red fixation cross for ~400 ms, followed by a dot array for ~200 ms, the fixation cross for ~400 ms, and a second dot array for ~200 ms. They were instructed to "get a rough idea of the sum" of these two arrays during the pause that followed for ~1300 ms. This followed by the word "TEST" for ~700 ms, signaling that the test array would appear next. The third "test" array then appeared for ~200 ms. The task was to determine whether the test array had fewer or more dots than the sum of the first two. The Crossmodal Addition case was similar, except that one of the addends was a tone sequence and one was a dot array. Half of the "test" sets were tone sequences; half were dot arrays. Responses were made by pressing one button for "fewer" and another button for "more." The terms fewer and more were used rather than smaller and bigger because the latter can be used to refer to mass nouns, while the former can only be used for count nouns; it is possible that subjects might be more inclined to make judgments based on non-numerical stimulus properties when judging smaller or bigger. Feedback was given throughout the experiment; participants heard a high beep when they chose the correct answer and a low beep when they chose the incorrect answer.

Results.

The top plot in Figure 1 shows mean accuracy for all four conditions. Accuracy scores for Visual Comparison, Crossmodal Comparison, Visual Addition, and Crossmodal Addition were 76%, 73%, 72%, and 74%, respectively. A one-way ANOVA demonstrated no accuracy differences among the conditions ($p > 0.2$). Reaction times are shown in the lower panel of

Figure 1. Mean RTs for Visual Comparison, Crossmodal Comparison, Visual Addition, and Crossmodal Addition were 890 ms, 1209 ms, 974 ms, and 992 ms, respectively. A one–way ANOVA ($p < .002$) followed by a Tukey HSD test revealed that the mean RT for Crossmodal Comparison was significantly different from Visual Comparison ($p < .01$) and Visual Addition ($p < .05$). Figure 2 shows accuracy and reaction time for all four conditions as functions of comparison ratio.

Discussion.

We found that our participants were indeed able to perform the addition task extremely well: there was no performance cost for Addition compared to Comparison, nor was there any performance cost for operations performed across modalities compared to those performed within modality. Accuracy also tended to decrease with larger comparison ratios, in accordance with Weber's Law. The finding that operations performed across different modalities are not more difficult than those performed within a modality is consistent with some previous results involving comparison tasks. (Barth, Kanwisher, and Spelke 2001) These studies provide additional evidence for the abstract nature of approximate representations of large numbers; adults can perform approximate arithmetic operations quite easily, even when the modalities of the addend sets differ. It is difficult to explain our puzzling finding that the more complex addition task, which itself contains the simpler comparison task, does not produce worse performance (as measured by accuracy or reaction time) than the comparison task alone.

## 3.2 Chapter 3 Experiment 2

In Experiment 2, we asked participants to perform a similar task to that in Exp. 1, but we substituted subtraction for addition. Subtraction with Arabic numerals is often more difficult than addition (Fuson 1984); if the approximate arithmetic abilities involved in the current tasks underlie human abilities at exact arithmetic with Arabic numerals, we might expect to find that approximate subtraction is more difficult than approximate addition. However, some explanations of the difficulty of subtraction attribute deficits to learned verbal strategies (Baroody 1984; Fuson 1984). If these explanations are accurate, there should not be a subtraction deficit for a large approximate task. A deficit would mean that the source of subtraction's difficulty relative to addition goes beyond learned arithmetic strategies.

Because pilot testing suggested that including both addition and subtraction tasks in a session leads to task-switching problems, we contrasted performance in Comparison tasks to performance in Subtraction tasks. We concentrate here on within-modality operations, so all of the conditions involved only visual arrays.

*Method*

Participants. Eleven participants between the ages of 18 and 35 participated in the study. All had normal or corrected-to-normal hearing and vision.

Apparatus. The experimental apparatus was the same as in Exp. 1.

Design. All participants received the same two stimulus conditions in counterbalanced order. Visual Comparison and Visual Subtraction trials were blocked; participants received two blocks of 62 trials of each condition, for a total of 124 trials per condition. Before the experimental blocks began, there were 5 practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block (Visual Comparison, Visual Subtraction). The Subtraction problems were reversed versions of the Addition problems of Exp. 1; for example, if an Addition problem was equivalent to "15 + 25 = (35) vs. 42," the corresponding Subtraction problem might be "35 – 15 = (25) vs. 30." Therefore the magnitudes that were involved in the solving of the Exp. 1 Addition problem and its Exp. 2 Subtraction version were the same. The quantities to be compared, however (Sum vs Test in Exp. 1 and Difference vs. Test in Exp. 2) were clearly not the same. For this reason, the Comparison condition of Exp. 2 did not use the same numerosities as the Comparison condition of Exp 1; rather, in Exp. 2 these Comparison numerosities were matched to the comparisons made in the subtraction problems. Overall, the Comparison condition of Exp. 2 involved smaller quantities than the Exp. 1 Comparison condition. As in Exp. 1, there were four comparison ratios: 0.75, 0.8, 0.83, and 0.86.

Stimuli. The dot arrays used in this experiment were the same as those used in Experiment 1.

Procedure. The Visual Comparison condition of Experiment 2 was the same as the Visual Comparison condition from Experiment 1, except that the compared numerosities were matched in this case to the Subtraction problems rather than to the Addition problems. The Subtraction condition was essentially the reverse of the Visual Addition condition of Experiment 1: participants first saw a small red fixation cross for ~400 ms, followed by a dot array, a pause of ~930 ms, and a second dot array. They were instructed to "get a rough idea of the difference" of these two arrays, and they were told that the first array would always be bigger so the difference would always be positive. These two first arrays were followed by the word "TEST" and a third array. The task was to determine whether the third "test" array had fewer or more dots than the difference of the first two. Responses were made by pressing one button for "fewer" and another button for "more." Feedback was given throughout the experiment; participants heard a high beep when they chose the correct answer and a low beep when they chose the incorrect answer.

Results.

The plot in the top panel of Figure 3 shows mean accuracy for both conditions. Mean accuracy for the Comparison task was 78%, and mean accuracy for the Subtraction task was 66%.. Reaction times are shown in the bottom panel of Figure 3; the mean RT for the Comparison task was 1017 ms, and the mean RT for the Subtraction task was 1122 ms. Figure 4 shows accuracy and reaction time for both conditions as functions of comparison ratio.

Discussion.

This experiment demonstrated that approximate subtraction is more difficult than comparison, unlike approximate addition for which no deficit was detected. As in the previous study, accuracy tended to decrease with larger comparison ratios in accordance with Weber's Law. In order to investigate possible causes of the subtraction difficulty, we also performed two additional versions of the study. To assess whether performance is better when subjects are queried about the subtrahend rather than the difference, the same subtraction problems were presented to a naïve group with slightly altered instructions. We found no effect of reordering the problems so that the subtrahend was the basis for comparison. The second revision of the study was carried out in order to assess whether a subtraction problem could be performed better when presented to subjects as an addition problem. Again the same problems were presented to a new group of participants, with instructions altered so that they were described in terms of addition, not subtraction. In the original subtraction version, participants were essentially asked, "Does the first minus the second equal the third?" In this new reworded addition version, participants were asked, "Does the third plus the second equal the first?" There was no effect of this manipulation either, though many subjects reported mentally reframing the problems in terms of subtraction.

Because a great deal of research has presented evidence that summation of contiguously presented quantities is automatic, it seems quite possible that the form of these problems triggers automatic summation, which would be likely interfere with a subtractive process. If this is the case, we might have observed a subtraction deficit that was a product of our experimental design, rather than being caused by the approximate subtraction process itself. Could adults perform better when the set to be subtracted is shown moving away from the larger set, or disappearing, as typically happens in analogous experiments with small children? Experiment 3 addressed these issues.

### 3.3 Chapter 3 Experiment 3

The previous studies suggested that the typical difficulty differences found between symbolic addition and subtraction were also found in nonsymbolic addition and subtraction carried out upon visual quantities. However, Exps. 1 and 2 did not allow for direct comparison of addition and subtraction. Addition and subtraction were not combined in a single study previously because of concerns about the difficulty of switching tasks within a single testing session. Ideally, one would obtain addition and subtraction data from each subject, but the complications of task switching are not trivial in this case. Our procedure for both types of operation consisted of the sequential presentation of two arrays; depending on the study, these arrays were either added or subtracted. Presenting one quantity and then a second is a natural way to represent summation, but perhaps a less natural way to represent subtraction. If it is true that summation often occurs automatically, it would make sense that the subtraction process would be hampered by the way the problems were presented. If both addition and subtraction were being performed in the same testing session, any such effected could be amplified.

If the subtraction deficit that we observed with nonsymbolic quantities was in fact due to the structure of the problem as it was presented, then it is possible that nonsymbolic subtraction is not really more difficult than nonsymbolic addition. Could a more natural representation of the subtraction operation – a more "ecologically valid" version of the task - improve subtraction performance relative to addition performance? To investigate this, we developed a nonsymbolic arithmetic paradigm patterned after the arithmetic studies performed on human infants and nonhuman primates (Wynn 1992; Hauser, MacNeilage, and Ware 1996; Sulkowski and Hauser 2001). Instead of simply presenting quantities and instructing the subjects to perform a particular operation, we used sequences of events that suggested those operations. For the addition task, we presented a first visual quantity, occluded it, and presented a second quantity that moved behind the occluder to join the first quantity. Then the occluder was removed to reveal the third quantity. Participants judged whether the third quantity was smaller or larger than the sum. For the subtraction task, the first quantity was presented and occluded, and the second quantity moved out from behind the occluder. Then the occluder was removed to reveal the third quantity, and participants judged whether it was smaller or larger than the difference. One group of participants performed these new versions of the addition and subtraction tasks, and another group performed the original tasks. We predicted that this new occlusion manipulation could improve subtraction performance relative to the original formulation of the task.

*Method*

Participants. Thirty-five subjects between the ages of 18 and 35 participated in the study. All had normal or corrected-to-normal hearing and vision. One of the 35 was excluded from the final analysis for performance more than 2 standard deviations below the mean.

Apparatus. The experiment was carried out using the same equipment and environment as the previous studies.

Design. Experiment 3 employed a mixed design, with the new "screen" manipulation as a between-subjects factor and the operation manipulation as a within-subjects factor. The first group of 17 participants comprised the No Screen group, and each of these 17 received Addition and Subtraction conditions. The remaining 17 subjects made up the Screen group and these 17 also received both Addition and Subtraction conditions. To minimize task switching requirements in the No Screen group, roughly half the subjects received all of the addition trials first (2 blocks) and the other half received all of the subtraction trials first. In the Screen case, subjects received two blocks of each operation in ABBA order, with the order of operations counterbalanced across subjects. The addition and subtraction problems were identical across groups (the same addition problems from Exp. 1 and the subtraction problems from Exp. 2).

Stimuli. The No Screen group used the dot arrays from Experiment 1, in which each array contained dots of the same size, but across arrays the dot sizes varied. The Screen condition, unlike the No Screen condition, required that the dot arrays be contained in small areas of the computer screen. Because of this, there was not enough space to vary the dot sizes in the same manner as in the no Screen condition (the Screen condition required smaller dot sizes overall). Instead, each array contained dots of varying sizes, ranging from 0.3 to 0.6 cm.

Procedure. The No Screen Addition condition involved the same additions that were used in Exp. 1, and the procedure was the same. This was the case as well for the no Screen

Subtraction condition and the problems of Exp. 2; for both operations, the first array was presented, followed by the second array, and then the Test array which the subject compared to the sum or to the remainder, depending on the operation to be performed. The procedure differed for the Screen group, though the actual addition and subtraction problems were identical, as were the discriminations to be made between the test arrays and the answers to those problems. Also, the timing was adjusted in order to keep the length of a Screen trial as similar as possible to the length of a No Screen trial. In the Screen Addition condition, the first addend array appeared in an imaginary rectangle in the bottom center portion of the display for ~400 ms. Then an opaque occluding rectangle appeared (the "screen"), concealing the array. After ~1000 ms, while the first array remained concealed, the second array appeared to travel from the top left portion of the display to the edge of the screen and then to disappear behind it. The motion of the second array was not smooth in order to prevent tracking of the array, which might lead observers to attempt to count the dots. Also, the second array disappeared behind the occluding screen in a single step so that no cues to numerosity were available from the gradual occlusion of the array. The second array appeared, traveled across the display, and disappeared in ~800 ms. Then after a pause of ~1300 ms, the screen disappeared, revealing the test array which remained for 400 ms. Subjects were told that they would see a representation of an addition problem; an initial quantity would be shown and covered, followed by the addition of a second quantity to the first, behind the screen. The screen would be removed, revealing a third quantity, which was an incorrect representation of the sum. The task was to determine whether the third revealed quantity was too small or too big to be the correct sum. The Screen Subtraction condition was analogous to Screen Addition but the quantities were the same as in the No Screen Subtraction case, and the second array appeared to move out from behind the concealing screen to the edge of the display rather than moving from the edge of the display toward the screen. In all conditions in Exp. 3, participants were *not* asked to make speeded responses, but were allowed to respond in whatever way seemed comfortable to them. They were, however, told that they would probably do better on the task if they chose the answer that fit their first impression, and did not think too long about the task.

Results.

The plot in Figure 5 shows mean accuracy for all four conditions. Accuracy scores were as follows: No Screen Addition 71%, No Screen Subtraction 66%, Screen Addition 75%, Screen Subtraction 70%. A mixed design ANOVA (Operation x Presentation) demonstrated a main effect of the within-subjects factor Operation ($p < .0005$) and a main effect of the between-subjects factor Presentation ( $p < .02$). There was no interaction. Figure 6 shows accuracy and reaction time as functions of comparison ratio for all four conditions.

Discussion.

In designing this study, we hypothesized that the manner of presentation used for our earlier addition and subtraction studies had been biasing our results against good subtraction performance. We predicted that the Screen condition might lead to the closing of the performance gap between the operations, in which case we would have seen an interaction between the factors (i.e., a subtraction deficit without the screen and no deficit with the screen) We did not observe any such interaction but we did see main effects of both factors. The subtraction deficit is clearly present in both the Screen and the No Screen groups. Interestingly, however, there is a clear effect of the mode of presentation as well – introducing the screen has improved performance on both the nonsymbolic addition task and the nonsymbolic subtraction task. Identical addition and subtraction problems produced higher levels of performance when they included simple animations that "acted out" the arithmetic operations. This was the case even though many of our participants were MIT students with sophisticated mathematical skills. When the trials were structured so that the operations to be performed were "acted out" by the elements of those operations, participants were more successful.

### 3.4 Chapter 3 Experiment 4

All of the previous experiments demonstrated that human adults can perform nonsymbolic addition and subtraction on large visual sets. Here, we inquire about adults' ability to perform multiplication and division as well. Patient studies have suggested that rote memorization is the basis for most multiplication, and have not shown evidence for preserved approximate abilities with multiplication when other broad quantity-based representations remain available (Dehaene and Cohen 1997). However, calculation of rates and ratios at some level seems to be a crucial part of animal foraging behavior. It is possible that adults possess approximate multiplication and division processes. However, the explicit multiplication and division of approximate quantity representations is more difficult to conceptualize than the addition and subtraction of such representations. While we can picture addition and subtraction as, for example, moves along a mental "number line" of sorts, it is not easy to come up with a comparable model for multiplication and division.

*Method*

Participants. Sixteen subjects between the ages of 18 and 35 participated in the study. All had normal or corrected-to-normal hearing and vision. Three of these participants were excluded from the final analysis; two of these did not complete all of the trials due to time constraints, and one failed to follow the experimenter's instructions.

Apparatus. The experimental apparatus was the same as in the previous studies.

Design. All participants received the same two stimulus conditions in counterbalanced order. Multiplication and Division trials were blocked; participants received two blocks of 62 trials of each condition, for a total of 124 trials per condition. Before the experimental blocks began, there were 5 practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block. There were 3 subdivisions within each condition so that different aspects of the problems could be matched across Multiplication and Division conditions. These subtypes were as follows:

Multiplication Type 1 (M1): operands of similar sizes, both relatively small but neither subitizable; comparisons matched to M2; inverse of D1

M1 example: $9 \bullet 6 = (54)$ vs. 36

Multiplication Type 2 (M2): operands of different sizes, multiplicand relatively large, multiplier subitizable; comparisons matched to M1; inverse of D2

M2 example: $18 \bullet 3 = (54)$ vs. 36

Multiplication Type 3 (M3): both operands relatively small, similar to M1 but some operands are subitizable; comparisons matched to D2

M3 example: $3 \bullet 6 = (18)$ vs. 12

Division Type 1 (D1): dividends large and matched to D2; divisors relatively small but not subitizable; comparisons matched to D3; inverse of M1

D1 example: $54 \div 6 = (9)$ vs. 6

Division Type 2 (D2): dividends large and matched to D1; divisors subitizable; comparisons matched to M3; inverse of M2

D2 example: 54 ÷ 3 = (18) vs. 12

Division Type 3 (D3): dividends smaller than D1 and D2; divisors subitizable; comparisons matched to D1

D3 example: 18 ÷ 2 = (9) vs. 6

This design allowed us to test various hypotheses about the steps taken in carrying out these operations. For example, the comparisons to be made in M1 and M2 are identical, but the operands are different. If nonsymbolic multiplication is performed through a series of sequential additions, reaction times should be greater for M1 than M2. As another example, M3 and D2 will allow us to assess performance differences across operations when the required comparisons are identical.

We hypothesized that multiplication and division performance would be worse than addition and subtraction performance, so the comparison ratios were slightly altered in Experiment 4. In the previous experiments, there were four comparison ratios: 0.75, 0.8, 0.83, and 0.86. In Experiment 4, the 2 most difficult ratios were dropped and one easier one was added, so the final ratios were 0.67, 0.75, and 0.8.

Stimuli. The dot arrays in this experiment were constructed in the same way as those in Experiments 1 and 2; there were no Screen conditions.

Procedure. The Multiplication and Division conditions in Exp. 4 were structured in the same way as the previous studies (except the Screen conditions from Exp. 3). The numerosities of the arrays, of course, were changed to form multiplication and division problems rather than addition and subtraction (as detailed in Design, above). As in Exp. 3, participants were *not* asked to make speeded responses, but were told that they would probably do better on the task if they chose the answer that fit their first impression.

<u>Results.</u>

The plot in the top panel of Figure 7 shows mean accuracy for both conditions. Mean accuracy for the Multiplication task was 67%, and mean accuracy for the Division task was also 67%. Reaction times are shown in the bottom panel of Figure 7; the mean RT for Multiplication was 1275 ms, and the mean RT for the Division task was 1368 ms. T-tests revealed that Multiplication and Division conditions were not significantly different either in accuracy or reaction time.

Figure 8 depicts the changes in performance with different comparison ratios for both conditions. Accuracy scores are shown on the top. A repeated measures ANOVA (Operation x Ratio) revealed a main effect of Ratio, $F(2,24) = 22.43$, $p < .0001$. Reaction times are shown in the bottom panel; ratio had no significant effect on reaction time.

Each condition in Exp. 4 (Multiplication or Division) was divided into 3 different subtypes. Plotted in Figure 9 are the paired comparisons of the subtypes in which the final comparison judgments are identical (M1 and M2, M3 and D2, D1 and D3). Accuracy is shown in the top panel. The first two pairs were not significantly different, but accuracy was higher for D3 than for D1 ($t(13) = -2.20$, $p < .05$). In the bottom panel, reaction times are plotted for each subtype; none of these pairs show significant RT differences.

Figure 10 shows the subtypes of the Multiplication and Division conditions which are inverses of each other (i.e., the division problems were created by inverting the multiplication problems). Accuracy is shown in the top panel, and reaction times in the bottom panel. M1 and

D1 show no significant differences in accuracy or reaction time. M2 and D2 also showed no differences either in accuracy or reaction time.

Discussion.

The results of Experiment 4 demonstrate that human adults can multiply and divide nonsymbolic visual sets, and that there are no performance differences between these operations as measured either by accuracy or reaction time, at least for the comparison ratios tested here. Though there were no overall group differences across these operations, most individual participants did report subjective differences that were reflected in their data. Often these subjective differences were explained in terms of strategies used to perform the operations. One participant reported that division seemed much easier to him than multiplication; he reported using a strategy he termed "caveman arithmetic" with division, which involved imagining, for example, 40 apples divided among 5 people. This participant reported that he had come up with no such strategy for multiplication and so his performance was worse. As in the case of all other operations addressed in this paper, participants reported that they made the correct choices when they judged "by instinct," and that conscious analysis of the problems tended to produce errors.

The final comparison ratios had a clear effect on both multiplication and division performance (see Figure 8); the most difficult ratio, 4:5, produced the worst accuracy scores. This result provides support for the idea that the participants were truly multiplying and dividing, for the comparison ratio could only have such an effect on performance if the products and quotients were in fact calculated and represented with some degree of accuracy.

We made a number of predictions about the results of the subtype comparisons. In the following section, examples of each subtype are included for clarification.

1. Is multiplication a separate process, or is it just sequential addition?

We predicted that the multiplication process would not be carried out by repeated summations; if multiple additions were used, we would expect reaction times to be greater for M1 [9 • 6 = (54) vs. 36] than M2 [18 • 3 = (54) vs. 36]. This would not necessarily be so if it took longer to assess the numerosity of the "18" array than the "9" array, but previous studies have shown the enumeration process does not seem to operate iteratively on visual sets (Barth, Kanwisher, and Spelke 2001).We can be confident that any reaction time difference between these two subtypes would not be due to the difficulty of the final comparisons, because these comparisons are matched (see Figure 9). Our results show that there was no difference in RT between these subtypes, suggesting that participants did not simply turn the multiplication operation into a series of additions.

2. Does the presence of a subitizable operand help?

When one of the operands in a problem is small enough that it can be subitized, it seems likely that performance would be improved compared to problems with no subitizable operands, because in the former problem type, at least one of the quantities would be known exactly. If this is the case, performance should be better for D2 [54 ÷ 3 = (18) vs. 12] than for D1 [54 ÷ 6 = (9) vs. 6]. The dividends are the same in these two subtypes, but the divisors are subitizable in D2 and they are not subitizable in D1. These two subtypes were not found to be significantly different in accuracy or reaction time (see Figure 9). However, D1 and D2 were necessarily *not* matched for the exact numerosities of the comparisons, so the effect of having a subitizable operand has not been perfectly isolated. The effect of operands that may be subitized can also be assessed by looking at differences between M1 [9 • 6 = (54) vs. 36] and M2 [18 • 3 = (54) vs. 36], which were already shown to be nonsignificant in (1). M1 and M2 do have perfectly matched comparisons, unlike D1 and D2, but their dividends are not identical as they were in D1 and D2 (again, see Figure 9). Therefore, again the effect of a subitizable operand is not perfectly isolated (which would of course be impossible, as the dividends and the quotients cannot be perfectly matched at once across sets of problems, unless the divisors are also identical). Taken together, these findings (that M1 vs. M2 and D1 vs. D2 did not produce performance differences) suggest that the presence of a subitizable operand, oddly, did not improve performance. This result is surprising not only because a subitizable operand has a numerosity that can be known exactly. It is also surprising because intuition would suggest that dividing a visually presented quantity into 2 or 3 parts, or multiplying it by 2 or 3, should be easier than dividing or multiplying by larger numbers. To divide a dot array by 2 or 3, for example, one only has to adjust one's perceptual grouping of the array so that it is not made up of one large group, but instead of 2 or 3 groups of roughly equal numerosities. Dividing the array into 6 such groups must be a more difficult feat of perceptual grouping; therefore, the similar performance levels found in subtypes with and without subitizable operands suggests that these operations are not being performed by the perceptual manipulation of the arrays.

3. When identical quantities are involved, is one operation more difficult than the other?

Two of the division subtypes were constructed by taking the inverses of the multiplication subtypes M1 [9 • 6 = (54) vs. 36] and M2 [18 • 3 = (54) vs. 36] . This means that the elements of the problems in D1 [54 ÷ 6 = (9) vs. 6] are the same as the elements in the M1 problems, albeit in a different order. The same goes for D2 and M2. The comparisons, of course, are not matched; they involve smaller numerosities for the division subtypes. Figure 10 shows accuracy and reaction time scores for these inverse pairs; there are no significant differences by either measure. This may mean that when the multiplication and division problems involve identical quantities, neither operation is more difficult. It is possible, however, that the differences in comparison numerosities affected this result in some way.

4. When final comparisons are perfectly matched, what kinds of problems are easier?

Three pairs of subtypes had comparisons that were identical to each other (Figure 9). Two of these pairs were matched within the same operation. The within-multiplication pair was M1 [9 • 6 = (54) vs. 36] and M2 [18 • 3 = (54) vs. 36], and the within-division pair was D1 [54 ÷ 6 = (9) vs. 6] and D3 [18 ÷ 2 = (9) vs. 6]. One of these pairs had comparisons that were matched across operations: M3 [3 • 6 = (18) vs. 12] and D2 [54 ÷ 3 = (18) vs. 12]. The only one of these three pairs that showed any performance differences was the within-division pair, D1 vs. D3. There was no RT difference, but accuracy was better for D3 than D1. This may be due to a combination of factors. The divisors were subitizable in D3 and not in D1, though this factor did not seem to affect performance in the pairs discussed earlier. The operand numerosities were also greater in D1 than D3; it is possible that the smaller quantities of D3 were easier to manipulate.

## 3.5 General Discussion

This set of experiments demonstrates that analog magnitude representations are available to human adults for calculation. First, we found that neither addition nor comparison of large sets is more difficult across stimuli that differ in modality and format. This finding is consistent with our previous results on comparison tasks (Barth, Kanwisher, and Spelke 2001), and it extends these results as well into the realm of calculation. Adults can perform approximate addition even when the modalities of the addend sets differ, supporting the claim that adults' processing of approximate numerosity is indeed based on truly abstract number representations. The fact that arithmetic operations can be performed overtly on nonsymbolic quantities *at all* is a novel result which deepens our understanding of the role that analog magnitude representations may play in human calculation. The notion that a biologically determined number sense, based on analog magnitudes, forms the basis for more complex human mathematical skills is consistent with these results.

Subtraction tasks with nonsymbolic visual sets are more difficult for adults than addition tasks. This result parallels the fact that symbolic subtraction is more difficult than symbolic addition, providing evidence against the claim that the symbolic subtraction deficit is due solely to learned verbal calculation strategies. The differences between subtraction and addition performance remained when we attempted to control for the possibility that subtraction was hampered by the involuntary, automatic addition of the quantities presented for subtraction. Thus, it seems that subtraction is a more difficult process than addition whether it is carried out through symbols or nonsymbolic quantities. Parallels such as this one between symbolic and nonsymbolic results support the idea that operations on analog magnitudes underlie learned operations on Arabic numerals. Experimental studies with infants and small children, and teaching methods used on small children, often involve "acting out" the processes that the researchers or teachers are trying to communicate. The present studies have shown that these enriched presentation methods, the "acting out" of addition and subtraction, can enhance performance even in mathematically advanced human adults.

According to studies in patients, rote memorization is the basis for multiplication processes. Even when patients retain the ability to use approximate quantity representations, they have not shown evidence of preserved approximate abilities with multiplication (Dehaene and Cohen 1997). However, these studies have dealt with the manipulation of Arabic numerals rather than nonsymbolic quantities, so it is possible that they failed to access approximate capabilities that did remain intact. We present evidence that both multiplication and division of approximate nonsymbolic quantities is possible though difficult. Though we have some preliminary evidence regarding the methods subjects may have used to carry out these calculations, it is difficult to come up with an intuitive way of multiplying a magnitude representation, for example, compared to adding such a representation. Further experimentation is needed before any conclusive statements can be made about the structure of approximate multiplication and division.

All of the operations tested in this set of studies demonstrated sensitivity to Weber's Law; performance was dependent on the ratio of the sum, difference, product, or quotient to the comparison numerosity. Figure 11 demonstrates this result; in the figure, accuracy scores were averaged across different experiments but within operations. This yielded an average percent correct for each type of arithmetic operation as a function of the comparison ratio. Though the basic accuracy means are obviously different for each type of operation, we see extremely regular patterns of performance, lending support to the idea that common representations, subject to the same rules, underlie all of these processes, and that the processes themselves add various sources of error which are reflected in performance. By demonstrating the relationship among these operations, and the relationships between some of these operations and their symbolic counterparts, these results contribute evidence to the idea that the noisy magnitude representations we share with other species are in fact available as operands in arithmetic computations, and that they may well may form the basis for more exclusively human symbolic numerical capabilities.

## Acknowledgments

Hilary Barth, Nancy Kanwisher, and Elizabeth Spelke, Department of Brain and Cognitive Sciences at the Massachusetts Institute of Technology and Department of Psychology at Harvard University.

This research was supported by National Institute of Health grant MH56037 to N. Kanwisher, National Institute of Health grant R37 HD23103 to E. Spelke, and National Science Foundation ROLE grant REC-0087721 to N. Kanwisher and E. Spelke.

## References


Ashcraft, M. H. (1992). "Cognitive arithmetic: a review of data and theory." Cognition **44**(1-2): 75-106.

Baroody, A. (1984). "Children's difficulty in subtraction: some causes and questions." Journal for Research in Mathematics Education **15**(3).

Barth, H., N. Kanwisher, and Spelke (2001). "The construction of large number representations in adults." under review.

Boysen, S. T. and G. G. Berntson (1989). "Numerical competence in a chimpanzee (Pan troglodytes)." J Comp Psychol **103**(1): 23-31.

Carey, S. (2001). Bridging the gap between cognition and developmental neuroscience: the example of number representation. Handbook of Developmental Cognitive Neuroscience. C. A. Nelson and M. Luciana, MIT Press: 415-431.

Church, R. M. and W. H. Meck (1984). The numerical attribute of stimuli. Animal Cognition. H. L. Roitblatt, T. G. Bever and H. S. Terrace. Hillsdale, NJ, Erlbaum: 445-464.

Dehaene (1997). The Number Sense. New York, Oxford University Press.

Dehaene, S. and L. Cohen (1997). "Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic." Cortex **33**(2): 219-50.

Dehaene, S., G. Dehaene-Lambertz, and L. Cohen (1998). "Abstract representations of numbers in the human and animal brain." Trends in Cognitive Sciences **21**(8): 355-361.

Dehaene, S., E. Spelke, P. Pinel, R. Stanescu, S. Tsivkin (1999). "Sources of mathematical thinking: behavioral and brain-imaging evidence [see comments]." Science **284**(5416): 970-4.

Fuson, K. (1984). "More complexities in subtraction." Journal for Research in Mathematics Education **15**(3).

Gallistel, C. R. (1990). The organization of learning. Cambridge, MA, US, MIT Press.

Gallistel, C. R. and R. Gelman (1992). "Preverbal and verbal counting and computation. Special Issue: Numerical cognition." Cognition **44**(1-2): 43-74.

Gibbon, J., C. Malapani, C. L. Dale, and C. R. Gallistel (1997). "Toward a neurobiology of temporal cognition: advances and challenges." Curr Opin Neurobiol **7**(2): 170-84.

Hauser, M. D., P. MacNeilage, and Ware (1996). "Numerical representations in primates." Proc Natl Acad Sci U S A **93**(4): 1514-7.

LeFevre, J. A., J. Bisanz, and L. Mrkonjic (1988). "Cognitive arithmetic: evidence for obligatory activation of arithmetic facts." Mem Cognit **16**(1): 45-53.

Meck, W. H., R. M. Church, and Gibbon (1985). "Temporal integration in duration and number discrimination." Journal of Experimental Psychology: Animal Behavior Processes **11**(4): 591-597.

Olthof, A., C. M. Iden, and W. A. Roberts (1997). "Judgements of ordinality and summation of number symbols by squirrel monkeys (Saimiri sciureus)." J Exp Psychol Anim Behav Process **23**(3): 325-39.

Olthof, A. and W. A. Roberts (2000). "Summation of symbols by pigeons (Columba livia): the importance of number and mass of reward items." J Comp Psychol **114**(2): 158-66.

Resnick, L. B. (1983). A developmental theory of number understanding. The Development of Mathematical Thinking. H. Ginsburg. New York, Academic Press: 109-151.

Rumbaugh, D. M., E. S. Savage-Rumbaugh, and J. L. Pate (1988). "Addendum to "Summation in the chimpanzee (Pan troglodytes)"." J Exp Psychol Anim Behav Process **14**(1): 118-20.

Rumbaugh, D. M., S. Savage-Rumbaugh, and M. T. Hegel (1987). "Summation in the chimpanzee (Pan troglodytes)." J Exp Psychol Anim Behav Process **13**(2): 107-15.

Stadler, M., D. Geary, and M. Hogan (2001). "Negative priming from activation of counting and addition knowledge." Psychological Research **65**: 24-27.

Sulkowski, G. M. and M. D. Hauser (2001). "Can rhesus monkeys spontaneously subtract?" Cognition **79**(3): 239-62.

Wynn, K. "Psychological foundations of number: numerical competence in human infants." Trends in Cognitive Science.

Wynn, K. (1992). "Addition and subtraction by human infants." Nature **358**: 749-750.

Xu, F. and E. S. Spelke (2000). "Large number discrimination in 6-month-old infants." Cognition **74**(1): B1-B11.

**Figure 1.** Accuracy and reaction time totals for the 4 conditions of Experiment 3.1 (Visual Comparison, Visual Addition, Crossmodal Comparison, Crossmodal Addition).

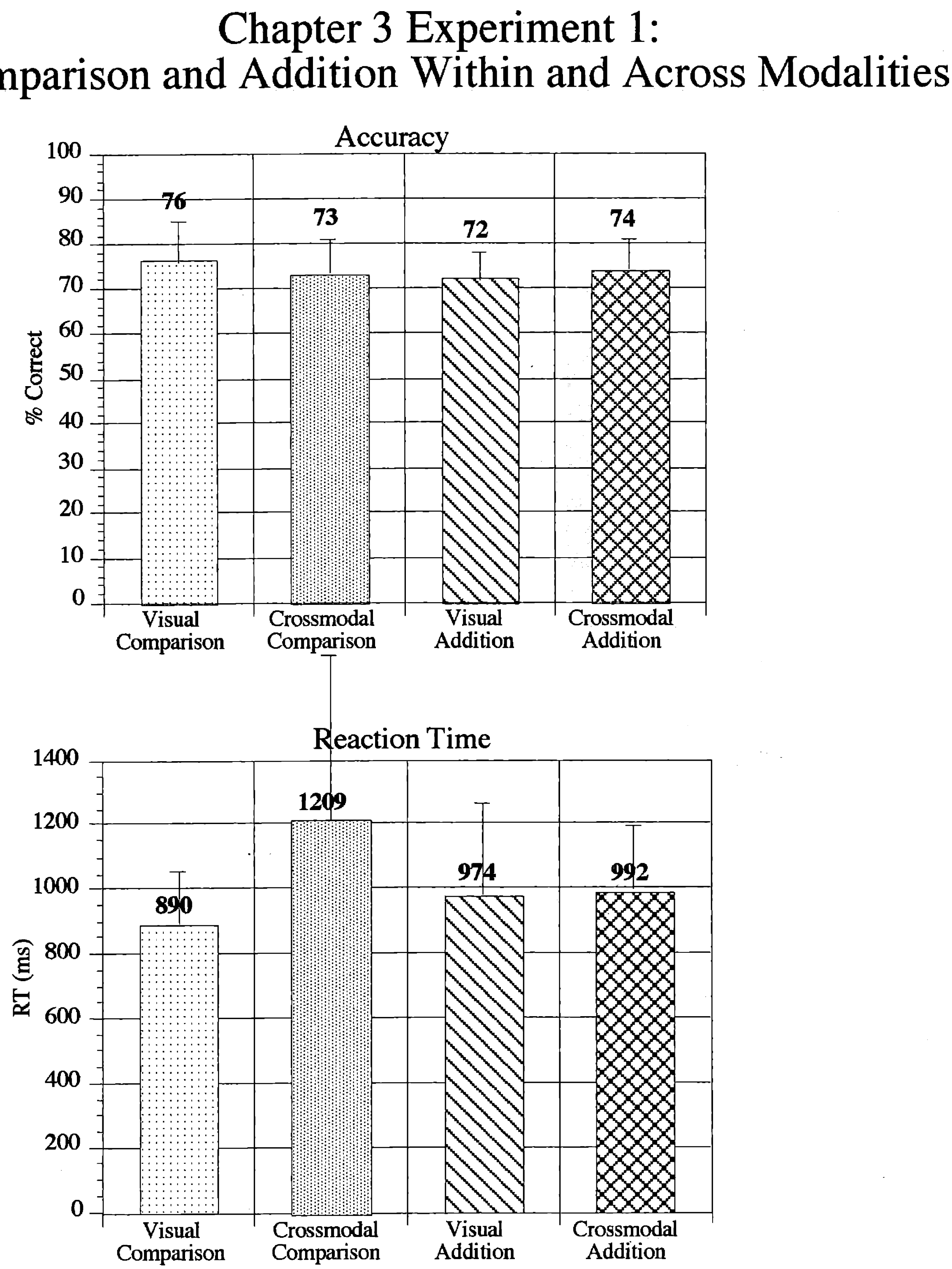


**Figure 2.** Accuracy and reaction time as functions of comparison ratio for Experiment 1.

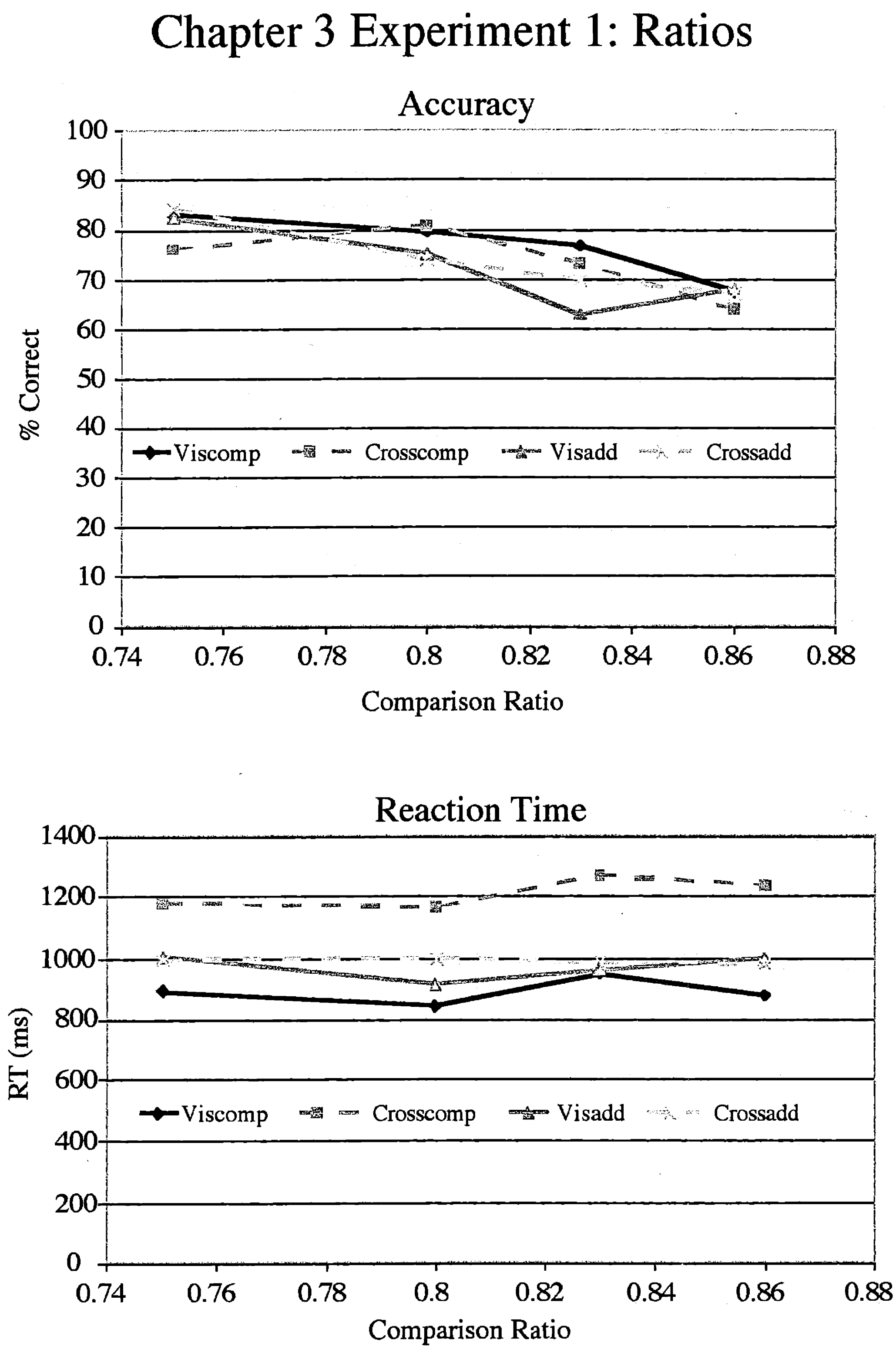


**Figure 3.** Accuracy and reaction time totals for the 2 conditions of Experiment 2 (Comparison and Subtraction, both visual).

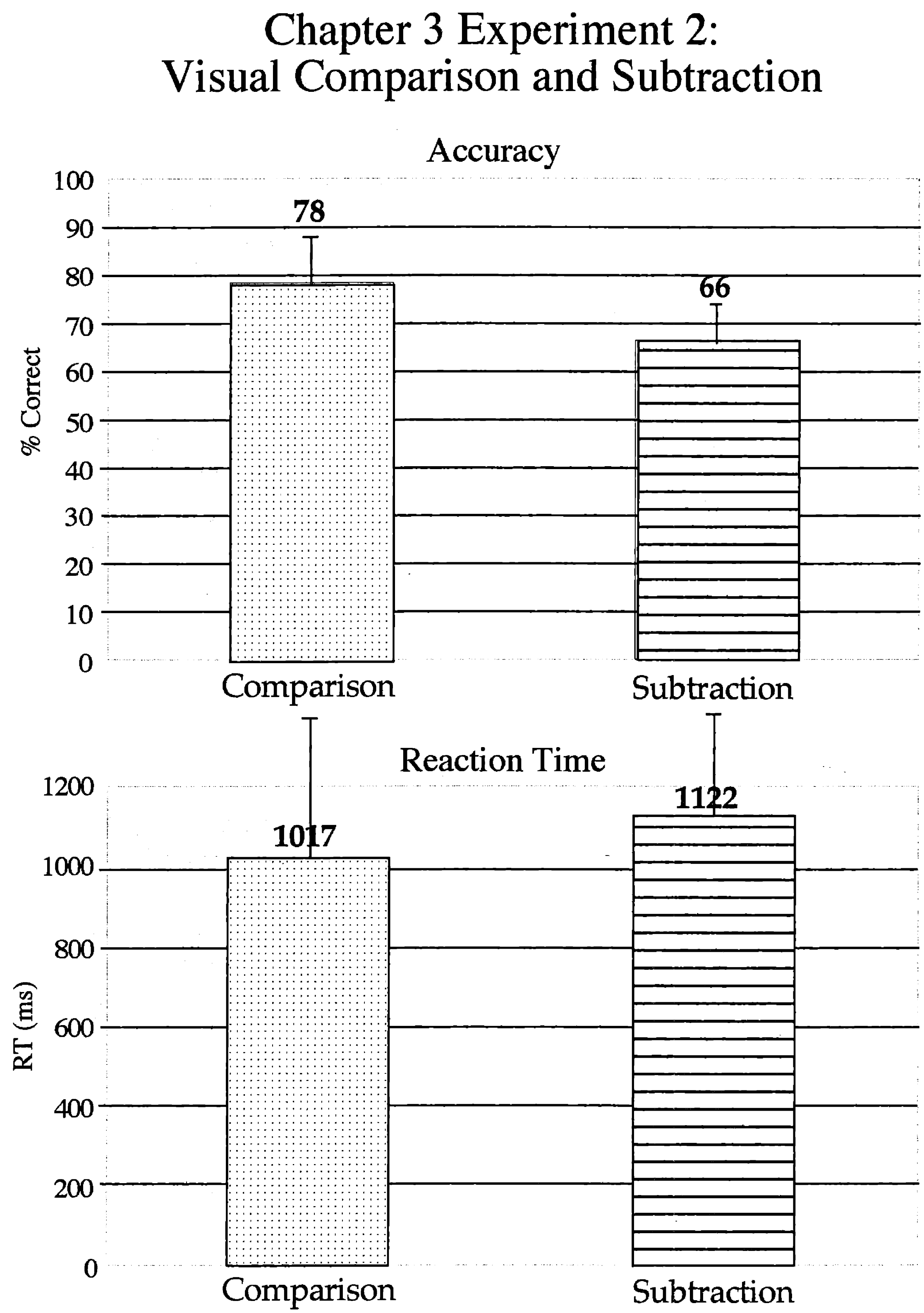


**Figure 4.** Accuracy and reaction time as functions of comparison ratio for Experiment 3.2.

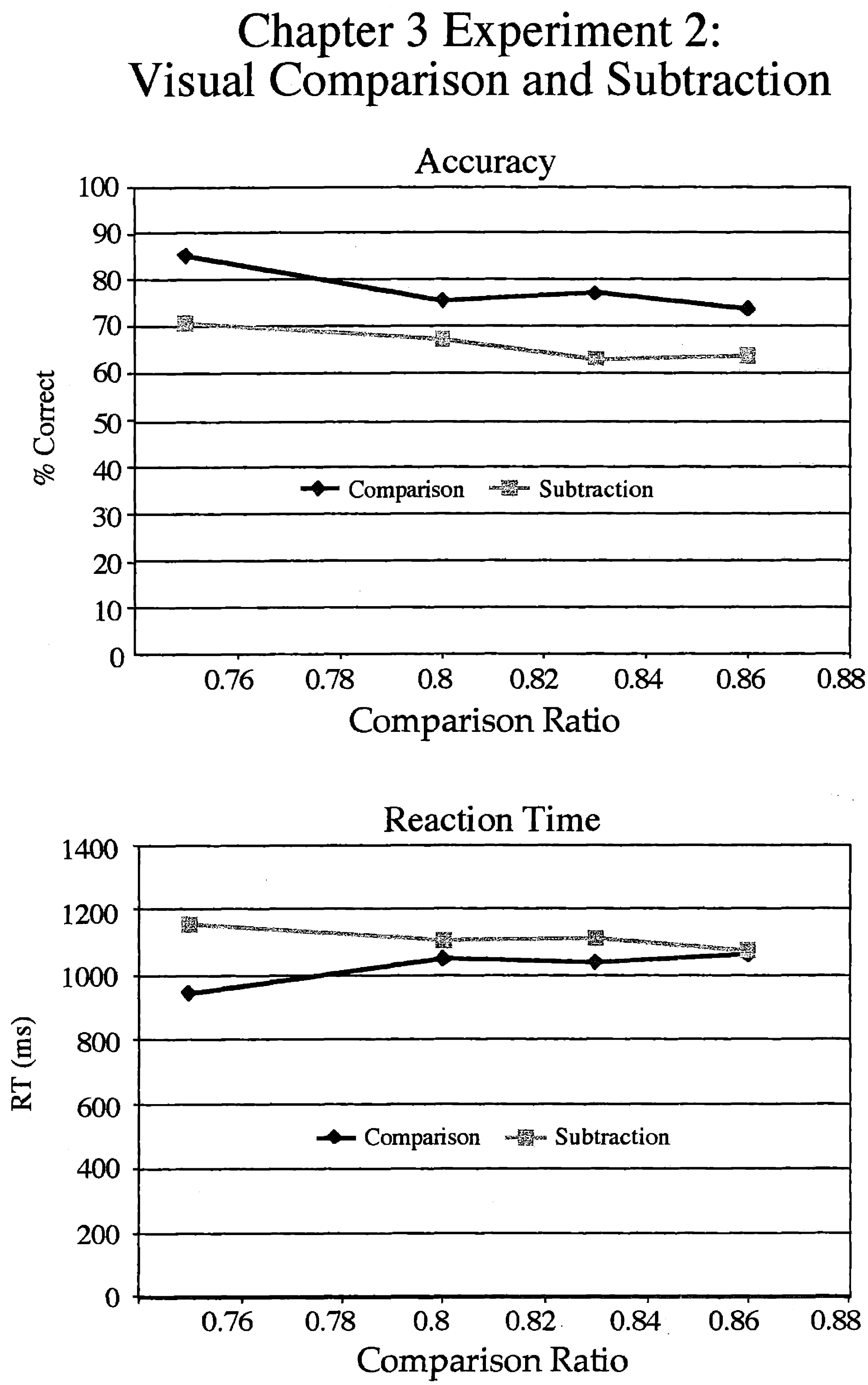


**Figure 5.** Accuracy and reaction time totals for the 4 conditions of Experiment 3.3 (No Screen Addition, No Screen Subtraction, Screen Addition, and Screen Subtraction). No Screen Addition and No Screen Subtraction conditions correspond to one group of 17 participants; a different group of 17 are represented in the Screen Addition and Screen Subtraction conditions.

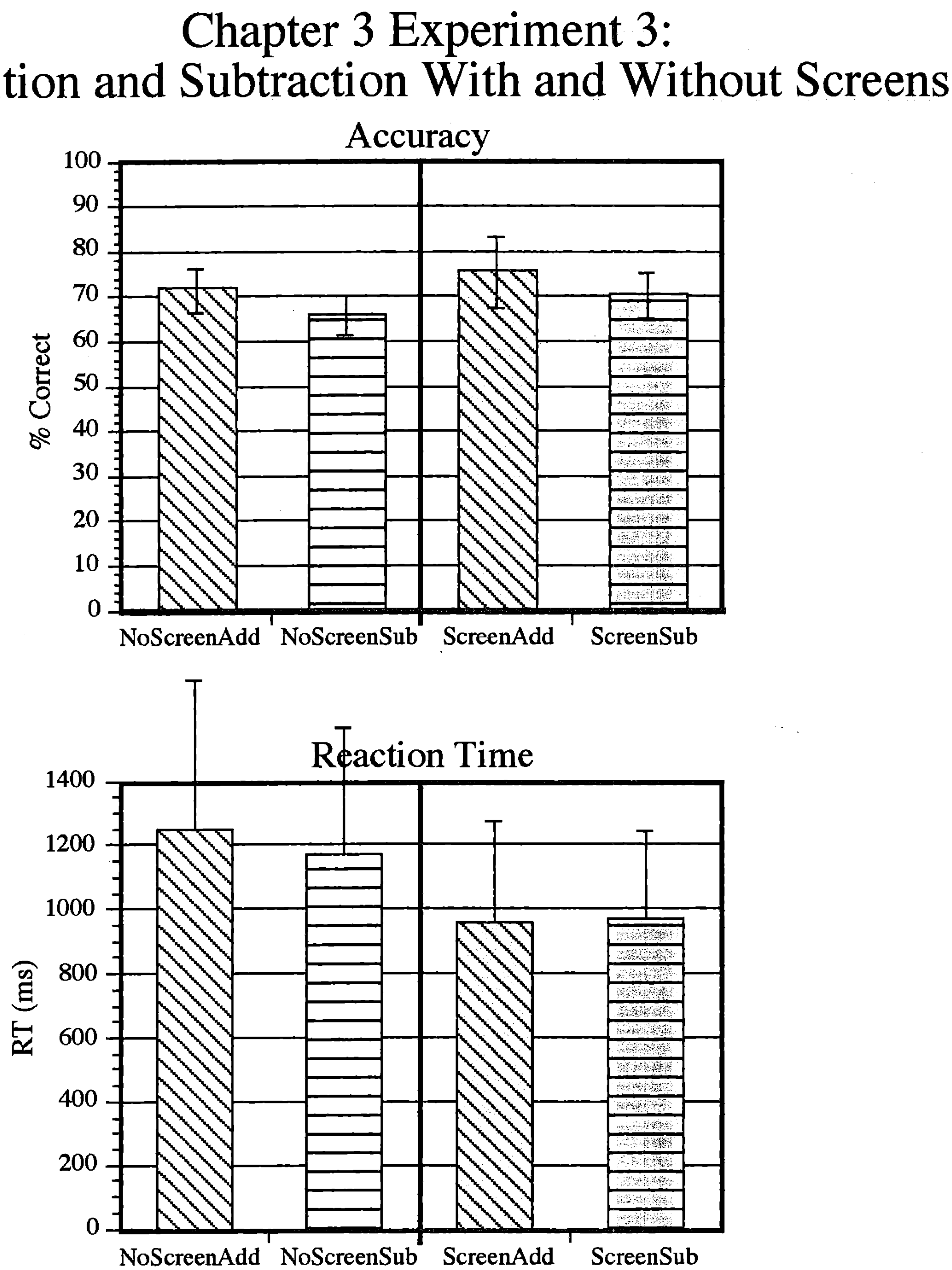


**Figure 6.** Accuracy and reaction time as functions of comparison ratio for Experiment 3.3

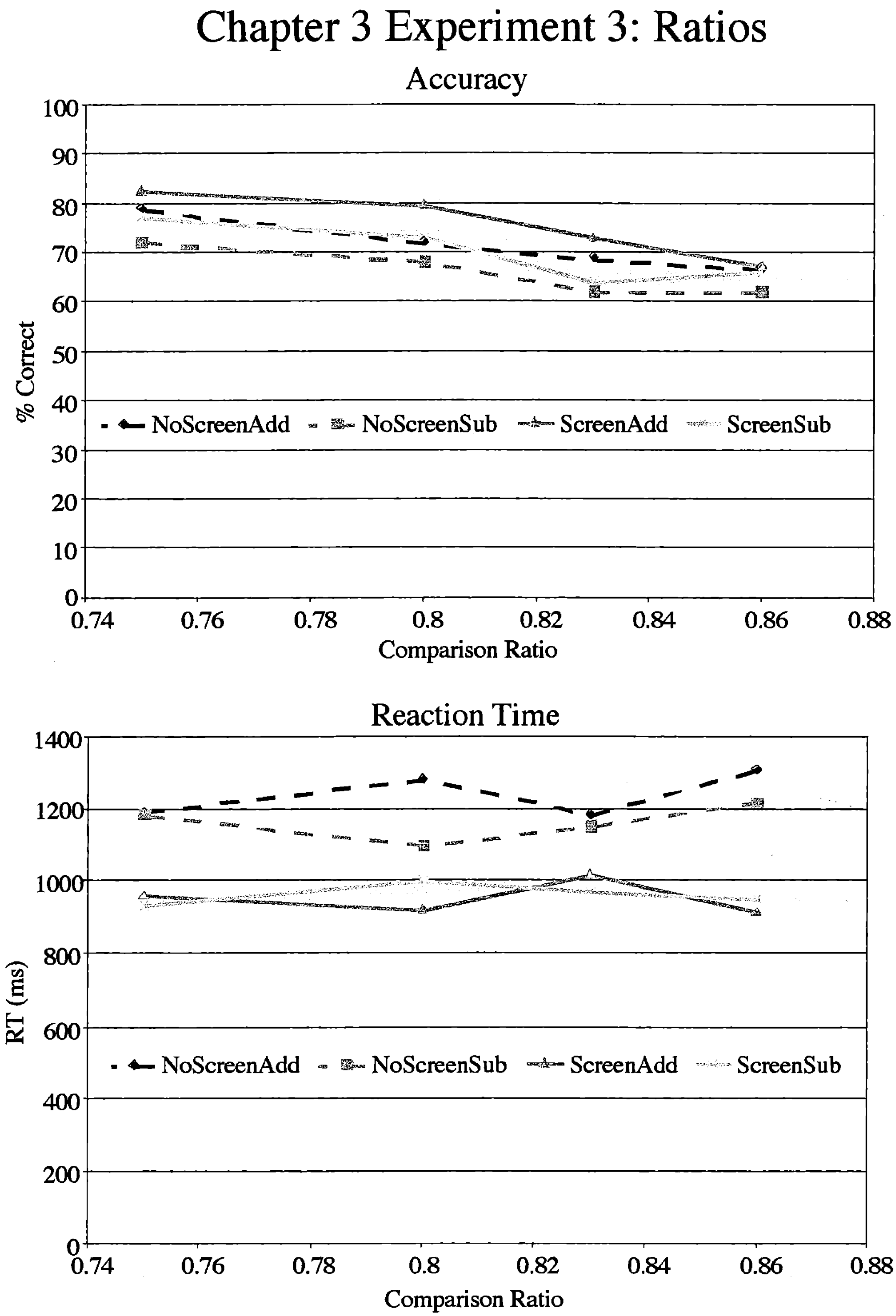


**Figure 7.** Accuracy and reaction time totals for the 2 conditions of Experiment 3.4 (Multiplication and Division, both visual).

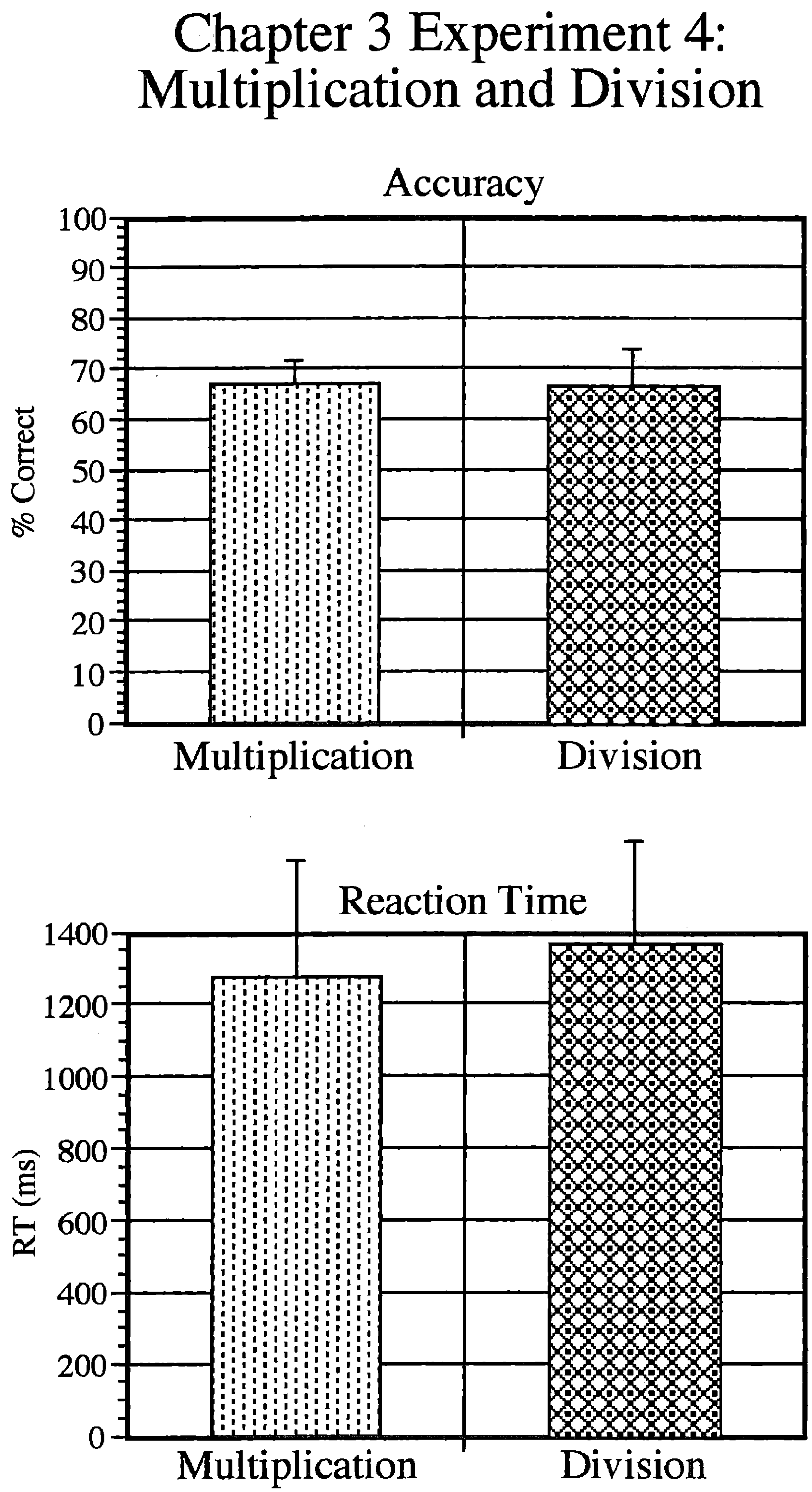


**Figure 8.** Accuracy and reaction time as functions of comparison ratio for Experiment 3.4.

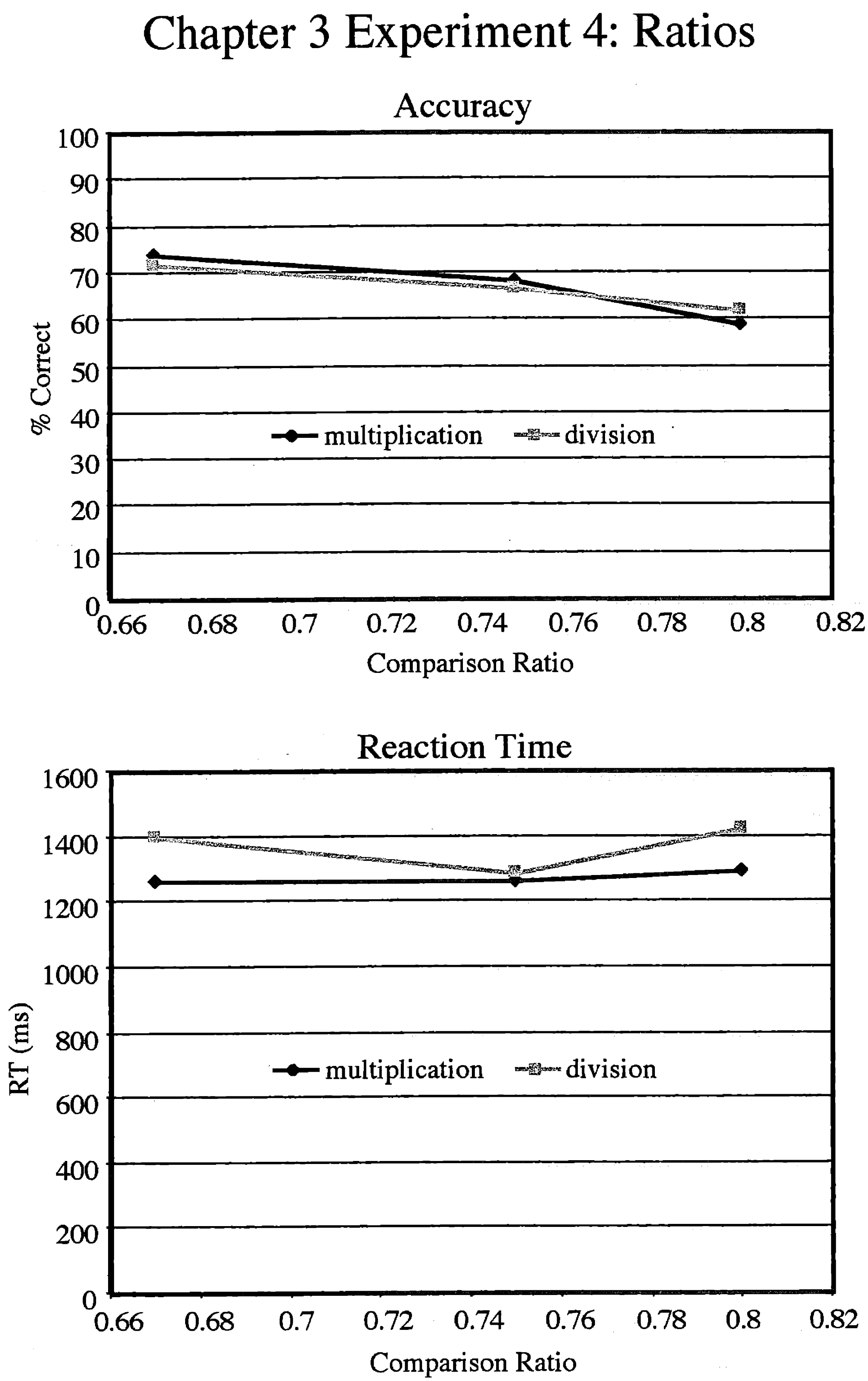


**Figure 9.** Accuracy and reaction time totals for Matched Comparisons in Experiment 3.4; shown are the pairs of subtypes across which the final comparisons were identical.

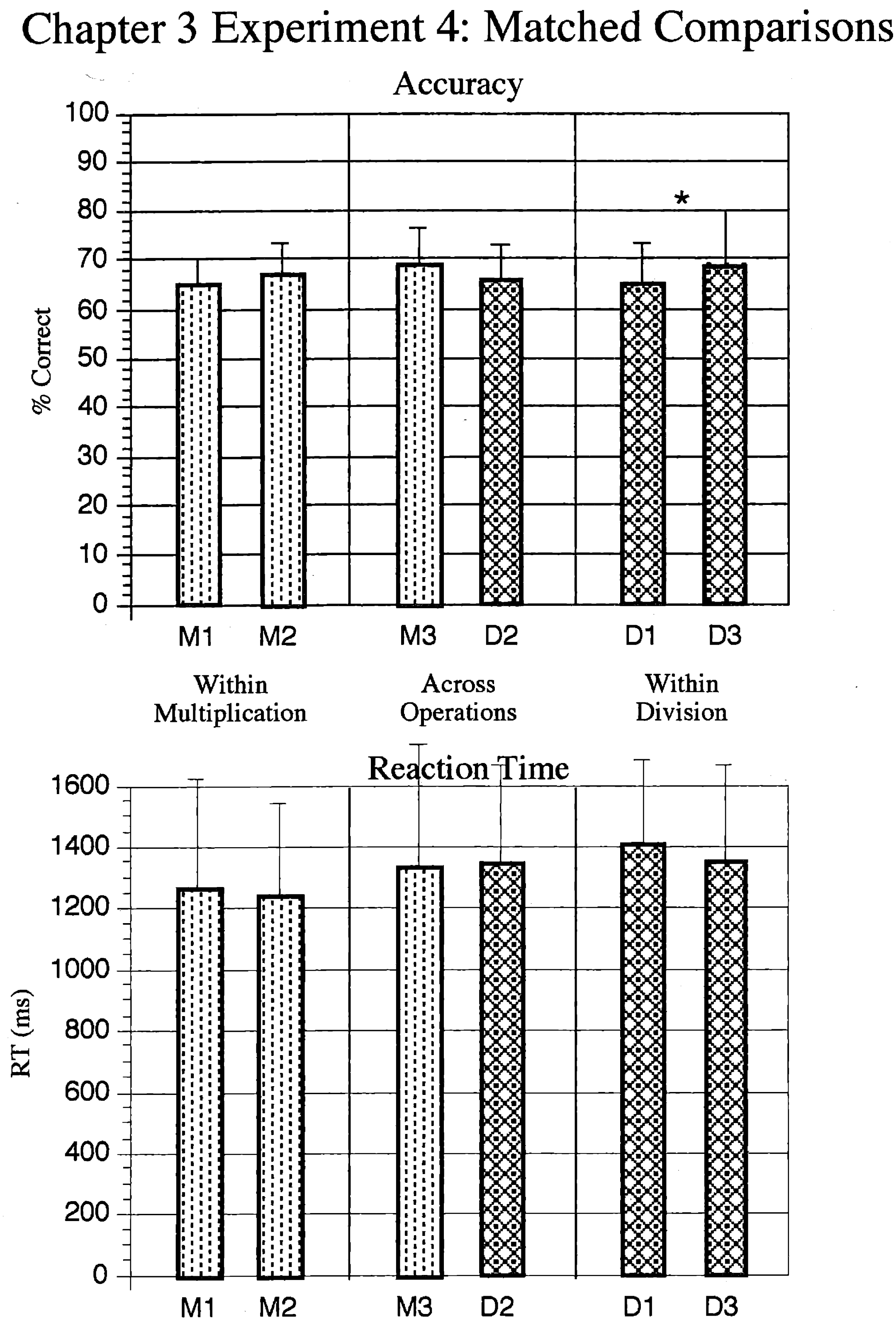


**Figure 10.** Accuracy and reaction time totals for Inverse Problems in Experiment 3.4; shown are the pairs of subtypes in which the division problems were created by inverting the multiplication problems.

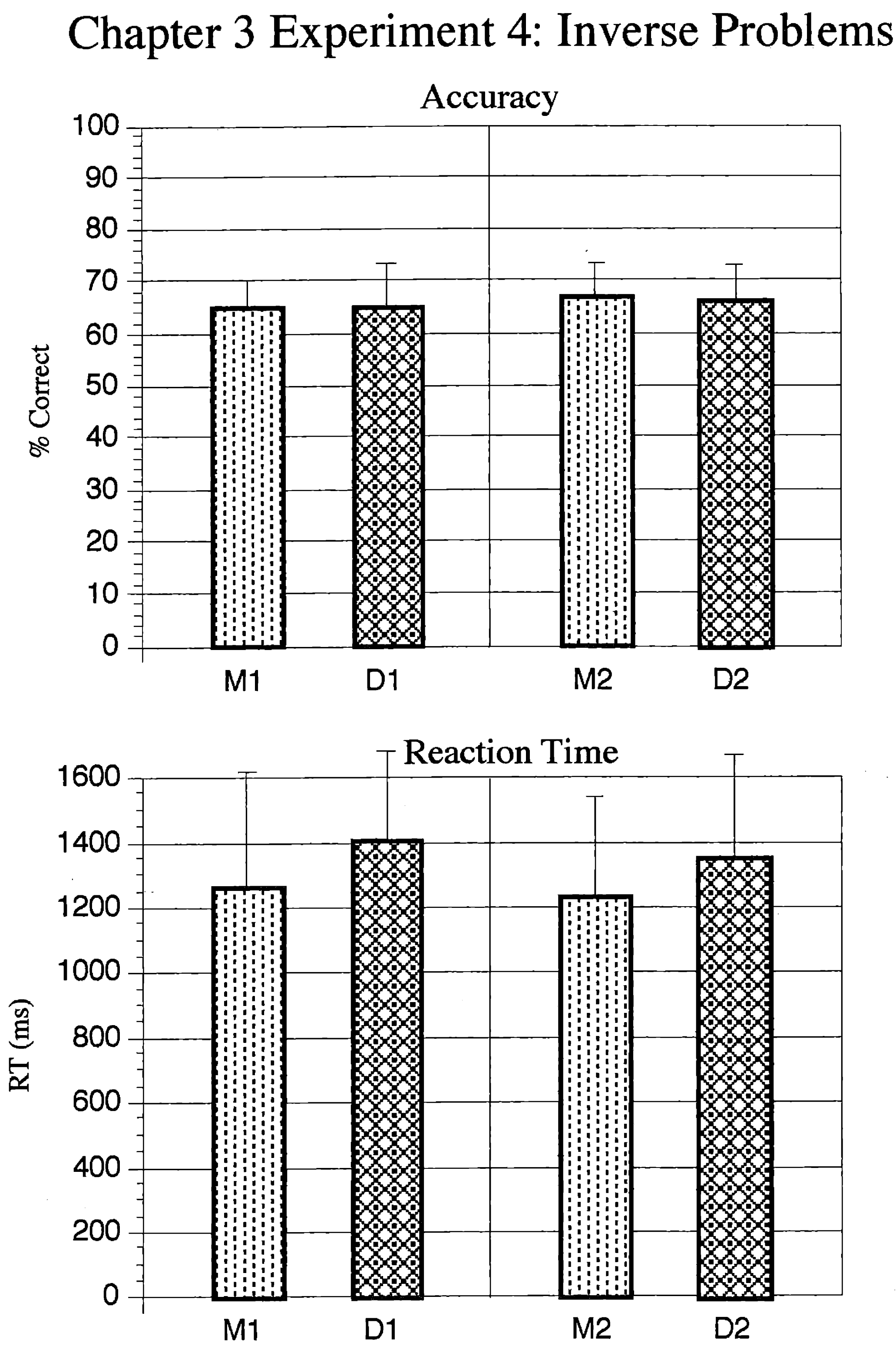
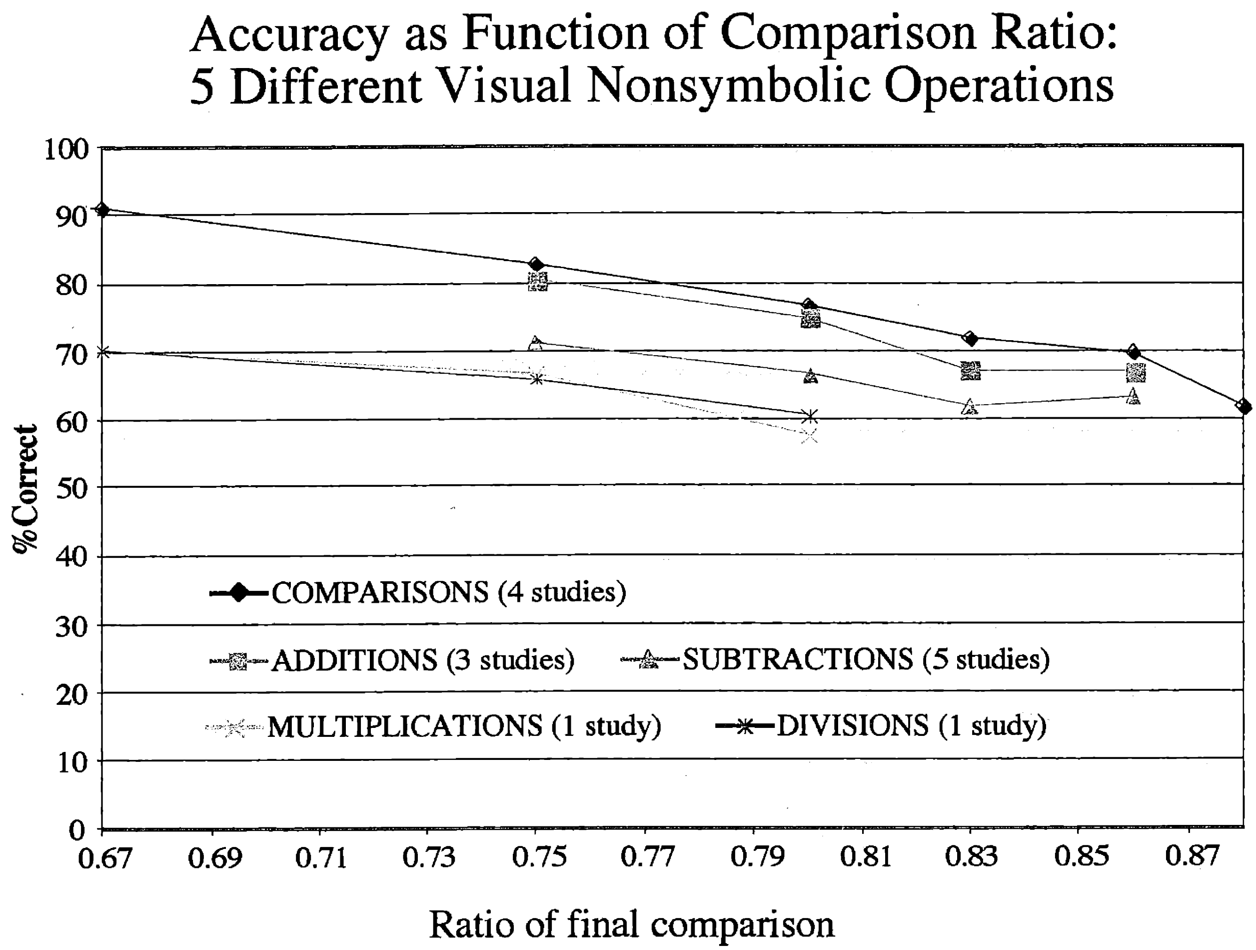


**Figure 11.** Accuracy as a function of comparison ratio for 5 different operations (comparison, addition, subtraction, multiplication, and division), averaged across multiple experiments.



# Chapter 4

# Conclusions

## 4.1 Abstract representation and enumeration mechanisms

These studies show first that adults' judgments of approximate numerosity are based on abstract representations of number by demonstrating that there is little or no cost for comparing or for adding numerosities across stimulus format or modality, relative to accuracy on intramodal and intra-format comparisons. Contrary to some theories of numerosity assessment, perceptual stimulus attributes cannot be directly responsible for numerosity judgments, suggesting that there must be some transformation of this perceptual information into an abstract form. Multiple perceptual cues may be combined to form abstract numerosity representations, perhaps analogous to the formation of depth percepts from binocular disparity, texture, and other cues.

The type of mechanism which constructs these representations has been further constrained by these experiments. When the difficulty of the numerical comparisons is held constant as defined by Weber ratio, adults take no longer to make comparisons between large visual sets than between small visual sets. This suggests that abstract numerosity representations are derived from perceptual representations by a non-iterative enumeration process of some sort. We believe that earlier findings of sensitivity to perceptual stimulus properties in enumeration tasks is due to the use of information that *is* tied to the perceptual properties of the stimulus in the construction of the numerosity representation, in a process analogous to depth perception as mentioned earlier.

Though enumeration mechanisms that operate iteratively have been invoked to explain many pieces of evidence from numerosity estimation tasks, our results are not compatible with such explanations at least for simultaneously present visual quantities. Our results do not contact the arguments for iterative enumeration mechanisms for sequential stimuli, and indeed this type of mechanism is intuitively suitable for such an enumeration task. Support for non-iterative enumeration for visual groups, however, has been found as well in human infants and small children (Xu and Spelke 2000; Huntley-Fenner 2001).

Chapter 2's results, taken together, suggest that human adults compare large discrete spatial quantities through the non-iterative construction of representations of numerosity, and that these representations are independent of the modality or format of the stimulus.

## 4.2 Calculation with nonsymbolic quantities

Chapter 3's experiments demonstrate that analog magnitude representations are available to human adults for calculation. We found that neither addition nor comparison of large sets is more difficult across stimuli that differ in modality and format, a finding that is consistent with our previous results on comparison tasks (Barth, Kanwisher, and Spelke 2001), and it extends these results as well into the realm of calculation. Adults can perform approximate addition even when the modalities of the addend sets differ, supporting the claim that adults' processing of approximate numerosity is indeed based on truly abstract number representations. The fact that arithmetic operations can be performed overtly on nonsymbolic quantities *at all* is a novel result which deepens our understanding of the role that analog magnitude representations may play in human calculation. We also find that subtraction is a more difficult process than addition whether it is carried out through symbols or nonsymbolic quantities. Parallels such as this one between symbolic and nonsymbolic results support the idea that operations on analog magnitudes underlie learned operations on Arabic numerals.

All of the operations tested in this set of studies demonstrated sensitivity to Weber's Law; performance was dependent on the ratio of the sum, difference, product, or quotient to the comparison numerosity. Figure 11 demonstrates this result; in the figure, accuracy scores were averaged across different experiments but within operations. This yielded an average percent correct for each type of arithmetic operation as a function of the comparison ratio. Though the basic accuracy means are obviously different for each type of operation, we see extremely regular patterns of performance, lending support to the idea that common representations, subject to the same rules, underlie all of these processes, and that the processes themselves add various sources of error which are reflected in performance. By demonstrating the relationship among these operations, and the relationships between some of these operations and their symbolic counterparts, these results contribute evidence to the idea that the noisy magnitude representations we share with other species are in fact available as operands in arithmetic computations, and that they may well may form the basis for more exclusively human symbolic numerical capabilities.

## 4.3 Numerical competence in animals: relationship with continuous quantity

*Sensitivity to number*

Some theorists are of the opinion that competence with "number" rather than continuous quantity should not readily be attributed to animals. A foraging animal wants the largest quantity of food presumably by volume, roughly, or mass, not by number. Number and continuous quantity tend to covary, and many researchers have suggested that animals use number only as a last resort, preferring to judge by continuous quantity. Why then be sensitive to number at all, if continuous quantity will do the job? Consider a potential foraging situation, taking into account the typical perceptual characteristics of the natural world. Occlusion is a constant in vision; each particular object in this hypothetical scene may well be only partially visible. Assessing foraging potential through continuous quantity in this situation will not necessarily yield an accurate representation of the food that's present, because continuous variables like visible area are affected by occlusion. However, if an animal is sensitive to discrete number, it can mark each visible segment as an object and assess the total amount of food accurately. This explanation for animals' sensitivity to large number accords with many other findings that nonverbal creatures are sensitive to discrete objects.

*Calculation*

Often, discussions of calculation abilities in animals urge caution in considering much of the relevant research, because in many cases there are confounds of number and continuous quantity. These confounds may not be as problematic as some claim, though, if the issue in question is animals' operational ability rather than their ability to represent discrete number. Analog magnitude representations are just that – analog – and it is quite possible that the magnitude representation of a continuous quantity such as volume is indistinguishable from the analog representation of discrete number. If this is the case, there is no reason to think that it would be possible to calculate with a magnitude representation that originated in a continuous stimulus, but impossible to calculate with a representation that originated in a discrete stimulus. Evidence of arithmetical processing of magnitude representations of area, volume, or rate, therefore, may be relevant to the issue of arithmetical processing of number representations.

## 4.4 The form of analog representations of magnitude – linear or logarithmic?

Analog magnitude representations, like representations of perceptual properties such as brightness or loudness, are subject to Weber's Law. A lengthy historical debate concerns the source of these Weber effects in perception; the issue remains controversial in the realm of numerical cognition as well. The numerical version of the debate asks whether the Weber's Law effects are due to *logarithmic compression* of the number representation (Dehaene and Mehler 1992), or to a representation that exhibits *scalar variability* (Gibbon 1977; Gallistel and Gelman 1992). According to the first hypothesis, Weber's Law results from a number representation that is compressed such that the difference between 5 and 10 is the same as the difference between 20 and 40. On this view, discriminability depends on the ratios between numbers rather than the absolute numerical distance because the relative *subjective* differences in the number representation itself depend on the ratios of the *objective* numbers. This type of explanation has traditionally been favored for Weber's Law in psychophysics research (Brannon, Wusthoff, Gallistel, and Gibbon 2001). The second hypothesis states that the number representation is linear in its relation to objective number, but its variability increases proportional to the objective number's mean, so that larger numbers have a more uncertain location in the representation (Gallistel and Gelman 1992). On this view, discriminability depends on the ratios between numbers rather than the absolute numerical distance because the uncertainty in the representations of larger numbers means that they may overlap. Indirect evidence from animals supports this theory because duration and number appear to be represented very similarly in (Meck and Church 1983), and duration appears not to be represented in a logarithmically compressed form (Gibbon and Church 1981).

The main difficulty in distinguishing between these theories is that in most situations, they both make the same prediction. However, in a few select cases this is not true. For example, the according to the logarithmic hypothesis, the difference between two subjective magnitudes depends on their ratio. The difference between any pair of magnitudes will be the same as the difference between any other pair if their ratios are the same. This is not the case according to the scalar variability hypothesis, which says that the differences between subjective magnitudes are determined by the absolute distances between objective magnitudes, and not by their ratios. Clearly arithmetic operations on this second type of representation are simpler to contemplate, which is another reason the hypothesis has been favored by some (Gallistel and Gelman 1992; Brannon, Wusthoff, Gallistel, and Gibbon 2001). It seems to be the case that analog magnitude representations are used for arithmetic operations, so a number representation with a morphology that is easily adapted to such operations should be preferred. Recent work in pigeons has presented further evidence for the scalar variability theory, by taking advantage of the distinction described above. Experimenters determined that pigeons' perceptions of differences between quantities were dependent upon absolute numerical difference as predicted by scalar variability, not ratio as predicted by logarithmic compression (Brannon, Wusthoff, Gallistel, and Gibbon 2001). However, it has been suggested that the observed behaviors were not actually based on differences, and further that the forms of internal representations of this sort may not be accessible to experimental techniques (Dehaene 2001). While it is possible that the former criticism is valid, the latter does not seem very robust considering the large amount of information that experiments have yielded so far regarding the nature of number representation. There is no clear logical motivation for declaring this particular distinction beyond the reach of experimentation.

Our studies on arithmetic in human adults can make contact with this debate from a number of angles. First, it seems that if we know certain operations can be performed on a representation, there is some reason to favor a hypothesis in which the format of the representation is well-suited to these operations. So if we know that nonsymbolic subtraction is performed in pigeons at all, a representation that allows for easy subtraction should be preferred over one that obscures the subtraction process. This is not direct experimental evidence, of course, but merely an intuitive guideline. Now, in this example involving pigeon subtraction, one would probably not imagine that the pigeons actually overtly attempt to subtract quantities. The computations involved are, presumably, not being carried out purposely by the pigeon. In the foraging literature, and indeed in most of the huge body of comparative psychology literature, it has been demonstrated repeatedly that at some level, animals' brains are capable of complex calculations involving integrations, probability, and so on. Therefore it could easily be argued that the fact that pigeons can subtract does *not* mean anything about the format of the number representation, because the manner in which calculations might be carried out upon that representation does not need to be transparent to the human observer; it simply needs to be manipulable by the brain. This is true, but consider a similar task given to a human subject who is directed to subtract two nonsymbolic quantities. In this case, presumably the subject is overtly attempting to subtract quantities. If this task could be achieved, it might provide slightly stronger support for scalar variability than the pigeon example, because a representation that is amenable to the subtraction operation, and hence a method of subtraction that is transparent to the human observer, would seem to be advantageous. Again, this is not direct evidence, but another intuitive guideline. We cannot assume much about the internal calculations taking place based on introspection about the process.

Direct evidence against the logarithmic compression hypothesis is what the present studies have provided. We have demonstrated that addition and subtraction of nonsymbolic quantities is indeed possible for human adults, and that addition is even strikingly easy in that it appears no more difficult than simple comparison. If the logarithmic hypothesis were correct, the subjective sum of two magnitudes would be equivalent to their objective product; for example, 20 dots plus 20 dots would equal roughly 400 dots, and addition performance would be atrocious

(Gallistel, personal communication). Similarly, in the case of subtraction, the difference of two magnitudes, according to the logarithmic hypothesis, should depend only on their ratio. These subtraction experiments largely involved problems with identical operand ratios, so the task should have been impossible if the number representation were logarithmically compressed. It is possible that some elaborate translation mechanism exists, allowing logarithmic number representations to be translated into linear form for calculation purposes, but because this is unparsimonious in the extreme, it seems more probable that the scalar variability hypothesis is the better explanation of Weber's Law in numerical processing.

The nature of numerical knowledge in human infants, human adults, and nonhuman animals has received extensive attention from researchers in recent years. These studies have suggested answers to some of the outstanding questions regarding the scope of numerical knowledge in human adults. The first set of experiments presents new evidence about the form of number representations and the ways in which we might construct them, and the second set second explores the kinds of processes that make use of these magnitude representations. Continuing work will advance our understanding of the "number concept" as found in human adults and the relationship of nonlinguistic number knowledge to the elaborate mathematical abilities humans have developed.

## References


Barth, H., N. Kanwisher, and E. Spelke (2001). "The construction of large number representations in adults." under review.

Brannon, E. M., C. J. Wusthoff, C. R. Gallistel, and J. Gibbon (2001). "Numerical subtraction in the pigeon: evidence for a linear subjective number scale." Psychol Sci **12**(3): 238-43.

Dehaene, S. (2001). "Subtracting pigeons: logarithmic or linear?" Psychol Sci **12**(3): 244-6; discussion 247.

Dehaene, S. and J. Mehler (1992). "Cross-linguistic regularities in the frequency of number words." Cognition **43**(1): 1-29.

Gallistel, C. R. and R. Gelman (1992). "Preverbal and verbal counting and computation. Special Issue: Numerical cognition." Cognition **44**(1-2): 43-74.

Gibbon, J. (1977). "Scalar expectancy theory and Weber's law in animal timing." Psychological Review **84**(3): 279-325.

Gibbon, J. and R. M. Church (1981). "Time left: Linear versus logarithmic subjective time." Journal of Experimental Psychology: Animal Behavior Processes **7**(2): 87-108.

Huntley-Fenner, G. (2001). "Children's understanding of number is similar to adults' and rats': numerical estimation by 5-7-year-olds." Cognition **78**(3): B27-40.

Meck, W. H. and R. M. Church (1983). "A mode control model of counting and timing processes." Journal of Experimental Psychology: Animal Behavior Processes **9**(3): 320-334.

Xu, F. and E. S. Spelke (2000). "Large number discrimination in 6-month-old infants." Cognition **74**(1): B1-B11.