ELSEVIER

Cognition 86 (2003) 201–221

COGNITION

www.elsevier.com/locate/cognit

# The construction of large number representations in adults

Hilary Barth*, Nancy Kanwisher, Elizabeth Spelke

*Massachusetts Institute of Technology, Harvard University, Cambridge, MA, USA*

Received 31 May 2000; received in revised form 4 July 2001; accepted 23 August 2002

**Abstract**

What is the nature of our mental representation of quantity? We find that human adults show no performance cost of comparing numerosities across vs. within visual and auditory stimulus sets, or across vs. 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. © 2002 MIT Published by Elsevier Science B.V. All rights reserved.

*Keywords*: Construction; Large number representations; Adults

## 1. 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

---

* Corresponding author. Laboratory for Developmental Studies, Harvard University, 33 Kirkland St., Cambridge, MA 02138, USA.

*E-mail address:* hilary@wjh.harvard (H. Barth).

PII: S0010-0277(02)00178-6

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 (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Gallistel & Gelman, 1992). 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 (Ginsburg, 1991; Ginsburg & Nicholls, 1988; Ginsburg & Pringle, 1988; Massaro, 1976). 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 & Tuulmets, 1991, 1993; Allik, Tuulmets, & Vos, 1991). 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 6–8-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 (Moore, Benenson, Reznick, & Peterson, 1987; Starkey, Spelke, & Gelman, 1983, 1990). However, a recent replication attempt found no preference in either direction (Mix, Levine, & 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, & Levine, 1996). While it is certainly possible that infants possess modality-independent numerical ability, to date evidence of this ability is equivocal.

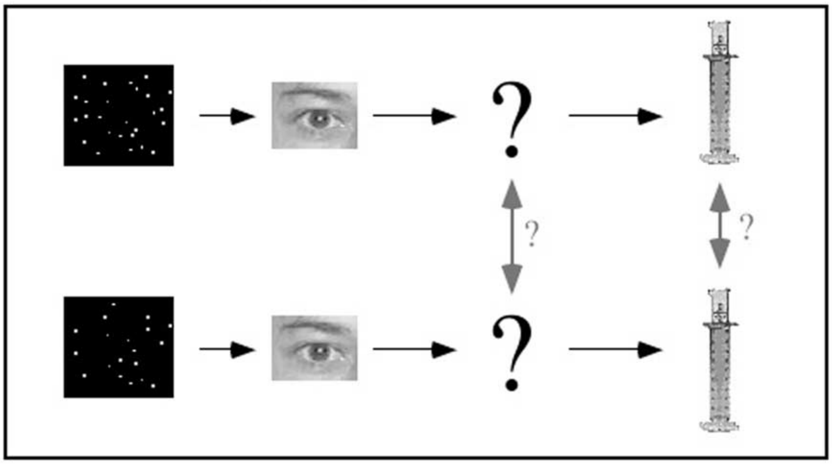
The animal literature provides far more robust evidence of the use of modality-independent numerical information. An impressive example of crossmodal transfer has been demonstrated in rats (Church & Meck, 1984). The animals were trained to press one lever when presented with two lights or white noise 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, & 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 & 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 (Fig. 1a). 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.

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 (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.

a.
Visual numerosity comparison:



b.
Crossmodal numerosity comparison:

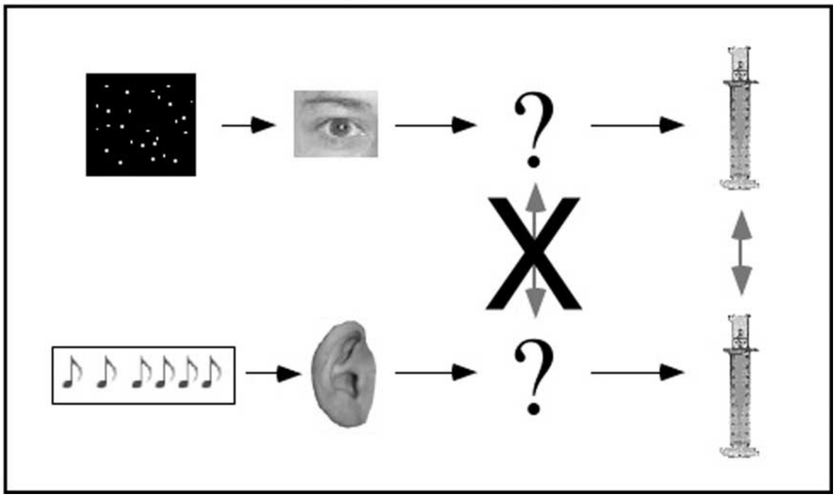


Fig. 1. (a) Schematic depiction of possible ways to perform a numerical comparison task within modality – such tasks cannot rule out the possibility that modality-specific representations are being compared. The question mark is an unknown perceptual representation of numerosity, which is specific to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

## 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”).

### 2.1. Method

#### 2.1.1. 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.

#### 2.1.2. 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.

#### 2.1.3. 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 ten 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, presented in these and reversed orders. “Same” and “Different” trials were equally frequent.

#### 2.1.4. 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. The tone presentation rate ranged from seven 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 to ~80 ms. The flash presentation rate ranged from six to nine flashes per second, also varying randomly so that the duration of the entire sequence was not a reliable cue to numerosity.

#### 2.1.5. 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.

### 2.1.6. Measures

We focus on the subjects' accuracy and not reaction time (RT). 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 crossmodal transfer of numerical information across modalities, then their performance on the crossmodal 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 crossmodal condition.

### 2.2. Results

Fig. 2 shows the mean accuracy for each condition in the left panel. The "worse unimodal" vs. Crossmodal comparison (see Section 2.1.6) is shown in the right-hand

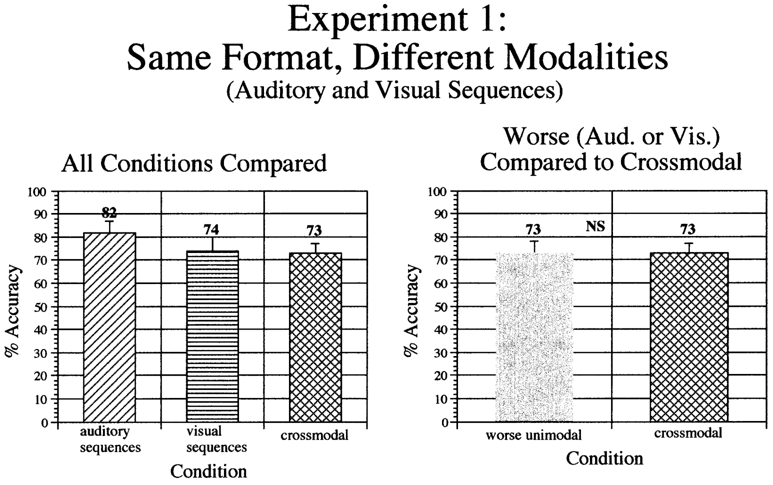


Fig. 2. Accuracy scores for Experiment 1.

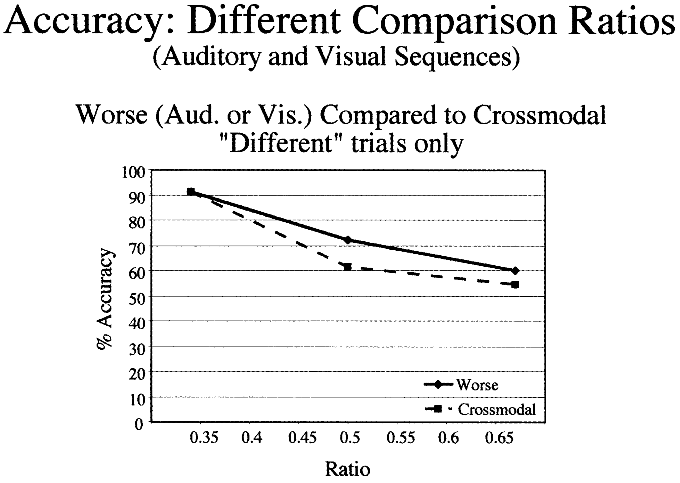


Fig. 3. Accuracy scores as a function of ratio for Experiment 1.

panel of Fig. 2. The "worse unimodal" and Crossmodal means were both 73%; these were clearly not significantly different ($t(14) = -0.03$, $P > 0.9$). Fig. 3 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 × Ratio) revealed a significant effect of Ratio ($P < 0.0005$); there was no effect of Condition.

### *2.3. 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[1] (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 Experiment 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 possi-

---

[1] 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 percussionist, 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.

bility by requiring subjects to make similar comparisons within a single modality but across stimulus presentation formats.

## 3. 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 of dots. Each subject therefore made numerical comparisons under three conditions: “Spatial”, “Temporal”, and “Crossformat”.

### *3.1. Method*

#### *3.1.1. 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.

#### *3.1.2. Design*

All participants received the following stimulus conditions in counterbalanced order. Spatial, Temporal, and Crossformat 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.

#### *3.1.3. 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 $\sim$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.

#### *3.1.4. 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 Crossformat trials between the two presentations in an attempt to equalize memory demands for Temporal and Crossformat conditions. The delay had a pseudorandomly selected duration ranging from that of the shortest visual sequence ($\sim$400 ms) to that of the longest ($\sim$2500 ms). The visual array appeared after this delay period.

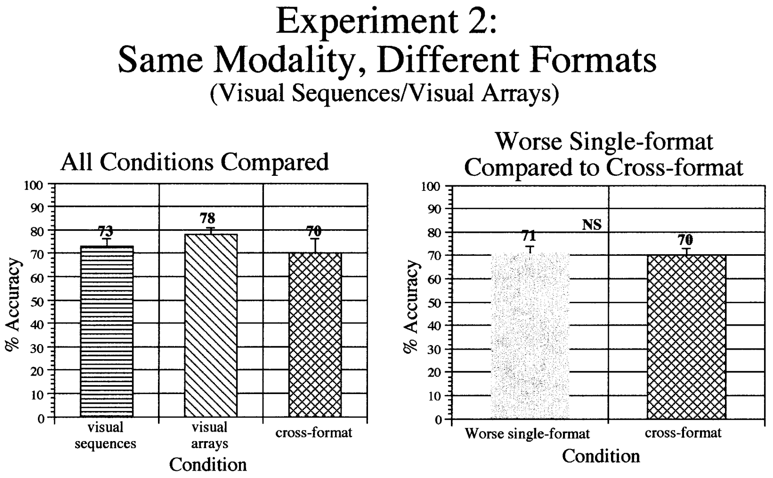


Fig. 4. Accuracy scores for Experiment 2.

### 3.2. Results

The left-hand plot in Fig. 4 shows the mean accuracy for each condition. As in Experiment 1, we determined the worse unimodal case for each participant individually (spatial or temporal) and compared the "worse unimodal" mean to the Crossmodal mean. This is shown in the right-hand panel of Fig. 4. The "worse unimodal" mean was 71% and the Crossmodal mean was 70%; these were not significantly different ($t(14) = 0.33$, $P > 0.7$). Fig. 5 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

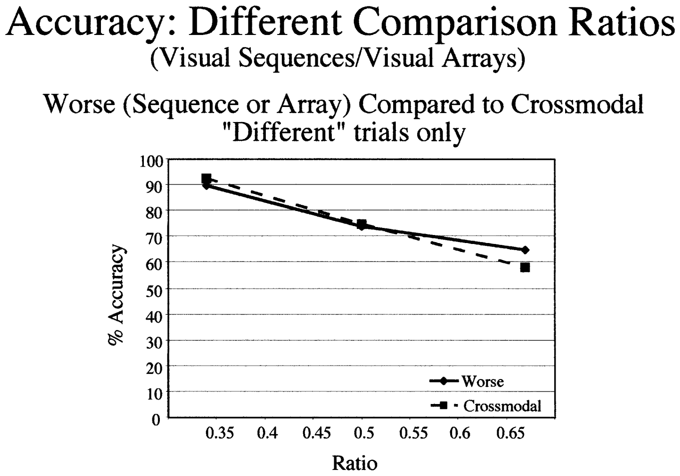


Fig. 5. Accuracy scores as a function of ratio for Experiment 2.

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 × Ratio) revealed a significant effect of Ratio ($P < 0.0005$); there was no effect of Condition.

### 3.3. 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.

## 4. 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/Crossformat”, shortened for ease to “Cross”).

### 4.1. Method

#### 4.1.1. 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.

#### 4.1.2. 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.

#### 4.1.3. Stimuli

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

#### 4.1.4. Procedure

The procedure was again the same as that used in Experiment 2, except for the replace-

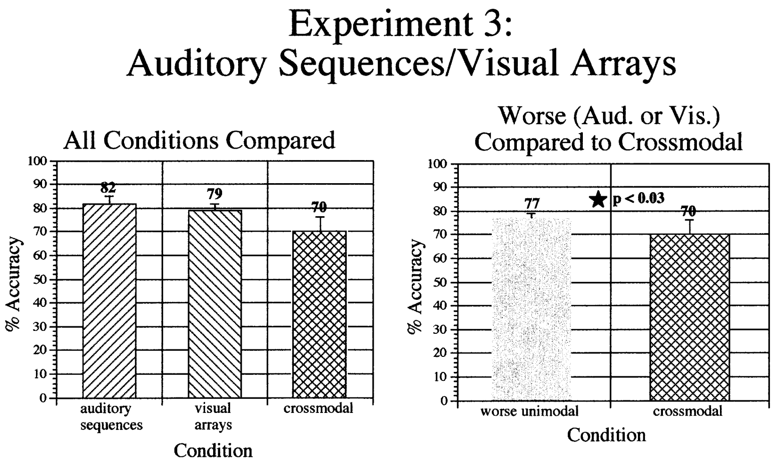


Fig. 6. Accuracy scores for Experiment 3.

ment of Experiment 2's visual sequences with visual arrays. A delay was introduced into the Cross trials as in Experiment 2.

### 4.2. Results

Fig. 6 shows the 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 or visual) and compared the Worse Uni mean to the Cross mean. This is shown in the right-hand panel of Fig. 6. The Worse Uni mean was 77% and the Cross mean was 70%; these were significantly different ($t(13) = 2.44$, $P < 0.03$).

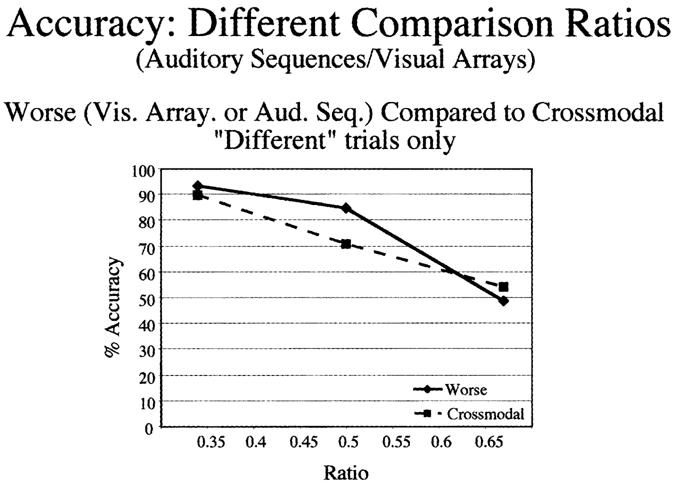


Fig. 7. Accuracy scores as a function of ratio for Experiment 3.

Fig. 7 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 × Ratio) revealed a significant effect of Ratio ($P < 0.0005$); there was no effect of Condition.

In this experiment, large differences in RT patterns across subjects appeared to reflect the adoption of very different strategies. Some subjects reported performing the cross-modal task by the use of a 1:1 correspondence strategy; these people described matching the tones to the dots as the auditory sequence played.[2] 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.

### 4.3. 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. Because there is a small difference between mean Worse Uni and Cross performance, there appears to be some cost for comparing numerosities in the Experiment 3 Cross condition. This pattern of results is somewhat different from those seen in Experiments 1 and 2. The crossmodal/crossformat 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/crossformat 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 (same vs. different) in our analysis, this conflict may contribute to the slight drop in performance across modalities and formats.

---

[2] 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".

It is also possible that alternative analyses to the "worse unimodal" test would produce different results. The "worse unimodal" test is a good way to measure crossmodal and/or crossformat costs under the assumption that crossmodal performance should be limited by the more difficult unimodal case. However, if crossmodal performance is dependent on some combination of the two unimodal cases, a comparison of mean unimodal performance to crossmodal performance would provide a better measure of crossmodal cost. This analysis revealed no significant difference for comparisons across modality (Experiment 1), and some cost for comparison across both modality and format at once (Experiment 3), as demonstrated by the "worse unimodal" test. The mean unimodal vs. crossmodal version of the analysis did reveal a small but significant difference for comparisons across format (Experiment 2), which may help to explain the source of the cost we find when comparing across both dimensions at once in Experiment 3. The conclusion remains that crossmodal/crossformat comparisons produce little or no performance deficit relative to within-modality or within-format comparisons.

## 5. 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 slightly 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 slightly greater difficulty observed within the crossmodal/crossformat 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 & Gelman, 1992), and those that are non-iterative (for example, sampling approximate density and area, distance between elements, or rate and duration; Church & 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 five different comparison ratios and four different absolute set sizes in order to assess the effects of set size and ratio on RT. 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 RT.

The other changes in Experiment 4 were introduced to meet possible objections to aspects of Experiments 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 crossformat comparisons were artificially masked by regression effects, yielding an overestimate of the difficulty of the unimodal conditions. In Experiment 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 Experiments 1–3.

## *5.1. Method*

### *5.1.1. 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 three experimental sessions.

### *5.1.2. Design*

All participants completed three separate experimental sessions, with conditions presented in counterbalanced order. Auditory/Temporal, Visual/Spatial, and Crossmodal/Crossformat (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 ten practice trials in each condition. At the beginning of each experimental block, participants were informed of the condition of the block. There were five possible comparison ratios, each presented in four absolute set sizes as shown in Fig. 8. In each block, each ratio/set size combination was presented twice (once in the order 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 |

Fig. 8. The numerosities used in Experiment 4.

### 5.1.3. Stimuli

Experiment 4 used the same basic stimuli as Experiment 3, visual arrays and auditory sequences, except that their numerosities were determined as shown in Fig. 8.

### 5.1.4. 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 to 60 ms within each sequence.

### 5.1.5. 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 three 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 each subject's first session, but used accuracy scores only from the last two sessions for the comparison. RT data, timed from the offset of the second stimulus, are also reported in this experiment.

## 5.2. Results

The mean Worse Uni accuracy score was 81%, and the mean Cross accuracy score was 77% (see Fig. 9). This slight accuracy difference between Cross and Worse Uni trials did not reach significance ($t(10) = -2.11$, $P > 0.05$). A repeated-measures ANOVA (Condition × Ratio × Group Size) revealed main effects of all three within-subjects factors (Condition: $F(2, 18) = 17.3$, $P < 0.0005$; Ratio: $F(4, 36) = 7.8$, $P < 0.0005$; Size:

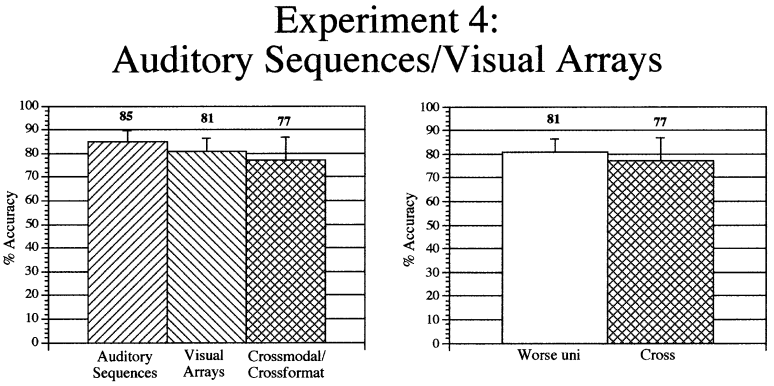


Fig. 9. Accuracy scores for Experiment 4.

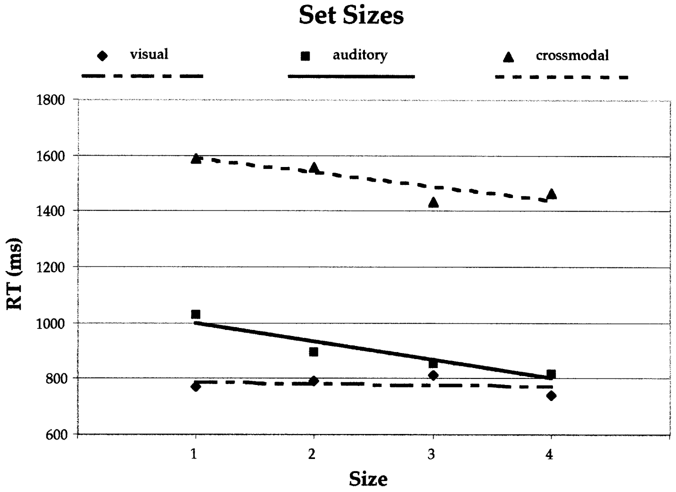


Fig. 10. Regression lines for RT as a function of set size in Experiment 4.

$F(3,27) = 3.8, P < 0.05$). The main effect of Size results from a *decreasing* linear trend in RT as set size increases, while RT *increases* with Ratio. Regressions are shown in Fig. 10 for Size and in Fig. 11 for Ratio.

### 5.3. Discussion

The key result of Experiment 4 is that response time does not increase at all with absolute set size; the comparison ratio alone determines the time necessary for these numerosity judgments. This finding is inconsistent with theories of numerosity estimation

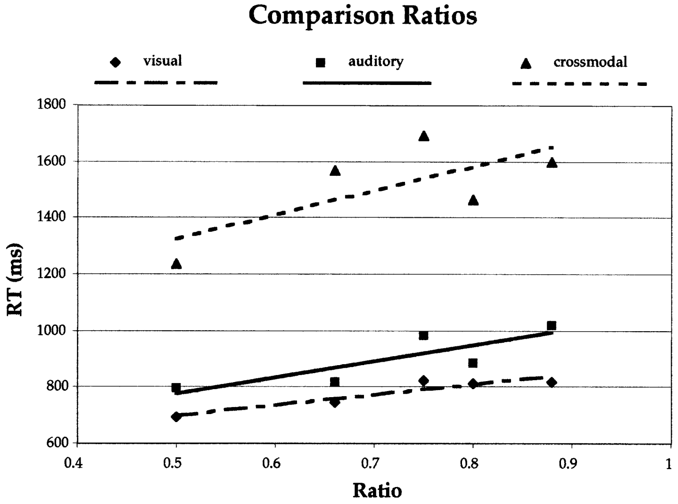


Fig. 11. Regression lines for RT as a function of ratio in Experiment 4.

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 RT, as discussed in the earlier experiments. Therefore, conclusions about enumeration mechanisms should not be drawn based solely on RT data from trials involving sequential stimuli. However, Experiment 4 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, which in fact show no increase in RT with increasing set size. Though it is possible that temporal/sequential sets are enumerated through a different process, this experiment can provide strong evidence of a non-iterative mechanism for the enumeration of spatial/simultaneous sets.

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 Experiment 4 (compared to Experiment 3), for several reasons. First, Experiment 4 included many more trials (240 per condition for each subject), so participants had more practice in Experiment 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.[3] A third possible explanation is the fact that the task was changed from "choose same or different" in Experiment 3 to "choose fewer or more" in Experiment 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.

There is another possible explanation which does not involve the actual comparison component of the task. All conditions were blocked in Experiments 1–3, so that the Crossmodal and Crossformat 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

---

[3] 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.

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 Experiment 4 is a robust result, and that it is not due to the same/different task of Experiment 3.

Experiment 4's results support our findings in Experiment 3 that crossmodal/crossformat comparisons are only slightly more difficult than intramodal or intraformat comparisons. There was a slight 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.

## 6. 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 intraformat comparisons, demonstrates that adults' judgments of approximate numerosity are likely to be 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 & Gelman, 1992), or protocounting (Davis & 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 RT. In the task used in Experiment 4, on the other hand, responses were limited to two alternatives, and we observed no increase in RT. A preverbal counting system could not have produced the patterns of RT we observed in Experiment 4; the fact that RT did not increase with set size shows that these sets were enumerated by a non-iterative process. Recent studies suggest that when young children estimate the numerosities of briefly presented groups of objects, they too produce patterns of performance consistent with the operation of non-iterative enumeration processes (Huntley-Fenner, 2001; Huntley-Fenner & Cannon, 2000).

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 & 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 & Gelman, 1992; Moyer & Landauer, 1967).

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: Kawai & Matsuzawa, 2000; Matsuzawa, 1985) 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.

## Acknowledgements

Our thanks to Mary C. Potter, Randy Gallistel, and two anonymous reviewers for their helpful comments. 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

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, *49*, 303–314.

Allik, J., & Tuulmets, T. (1993). Perceived numerosity of spatiotemporal events. *Perception & Psychophysics*, *53*, 450–459.

Allik, J., Tuulmets, T., & Vos, P. G. (1991). Size invariance in visual number discrimination. *Psychological Research*, *53*, 290–295.

Church, R. M., & Broadbent, H. A. (1990). Alternative representations of time, number, and rate. *Cognition*, *37*, 55–81.

Church, R. M., & Meck, W. H. (1984). The numerical attribute of stimuli. In H. L. Roitblatt, T. G. Bever & H. S. Terrace (Eds.), *Animal cognition*. Hillsdale, NJ: Erlbaum.

Davis, H. (1993). Numerical competence in animals: life beyond Clever Hans. In S. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: animal and human models*. Hillsdale, NJ: Erlbaum.

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

Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience*, *21*, 355–361.

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

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. Special issue: numerical cognition. *Cognition*, *44*, 43–74.

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

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

Ginsburg, N., & Pringle, L. (1988). Haptic numerosity perception: effect of item arrangement. *American Journal of Psychology*, *101*, 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*, B27–B40.

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

Kawai, N., & Matsuzawa, T. (2000). Numerical memory span in a chimpanzee. *Nature*, *403*, 39–40.

Lechelt, E. C. (1975). Temporal numerosity discrimination: intermodal comparisons revisited. *British Journal of Psychology*, *66*, 101–108.

Massaro, D. W. (1976). Perceiving and counting sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 337–346.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, *315*, 57–59.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (1996). Do preschool children recognize auditory-visual numerical correspondences? *Child Development*, *67*, 1592–1608.

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1997). Numerical abstraction in infants: another look. *Developmental Psychology*, *33*, 423–428.

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

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520.

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

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, *222*, 179–181.

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36* (2), 97–127.