Spontaneous supra-modal encoding of number in the infant brain

Giulia Gennari, Stanislas Dehaene, Chanel Valera, Ghislaine Dehaene-Lambertz

Spontaneous supra-modal encoding of number in the sleeping infant brain

Authors: Giulia Gennari[1,3]*, Stanislas Dehaene[1,2], Chanel Valera[1], Ghislaine Dehaene-Lambertz[1]*

[1] Cognitive Neuroimaging Unit U992, Institut National de la Santé et de la Recherche Médicale, Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Direction de la Recherche Fondamentale/Institut Joliot, Centre National de la Recherche Scientifique ERL9003, NeuroSpin Center, Université Paris-Saclay, 91191 Gif-sur-Yvette, France

[2] Collège de France, Université Paris Sciences Lettres (PSL), 75005 Paris, France

[3] Lead contact

*Correspondence: giulia.gennari1991@gmail.com; ghislaine.dehaene@cea.fr

Gennari, Giulia, Stanislas Dehaene, Chanel Valera, et Ghislaine Dehaene-Lambertz. « Spontaneous Supra-Modal Encoding of Number in the Infant Brain ». *Current Biology*, 17 avril 2023. https://doi.org/10.1016/j.cub.2023.03.062.

SUMMARY

The core knowledge hypothesis postulates that infants automatically analyze their environment along abstract dimensions, including number. According to this view, numerosity should be encoded quickly, pre-attentively, and in a supra-modal manner by the infant brain. Here, we provide a direct test of this idea by submitting the neural responses of sleeping 3-month-old infants, measured with high-density electroencephalography (EEG), to decoders designed to disentangle numerical and non-numerical information. The results show the emergence, in approximately 400ms, of a decodable number representation, independent of physical parameters, that separates auditory sequences of 4 versus 12 tones and generalizes to visual arrays of 4 versus 12 objects. Thus, the infant brain contains a number code that transcends sensory modality, sequential or simultaneous presentation, and arousal state.

Key words: infant, number, numerosity, EEG, decoding, MVPA

INTRODUCTION

Young animals, including human infants, are exposed to a complex physical world where survival depends on a fast and accurate understanding of the environment. Although current advances in artificial intelligence show that many structural features can be discovered from raw data, this comes at the cost of time, memory, and a large training set. Thus, evolution may have internalized, in the developing brain, strong priors about the most useful and reproducible physical laws. Some of these priors, shared by many species facing the same difficulties, constitute an initial core of knowledge [1] that structures perception and guides learning.

Number discrimination has been reported for mammals, birds, fishes and even insects, pointing to an ancient component of core knowledge which may have evolved independently in multiple lineages [2,3]. This capacity is only approximate: it does not support exact counting, but allows animals to distinguish sets of objects separated by a certain numerical ratio. Concerning humans, psychophysical assessments in adults have shown that approximate numerical judgments are susceptible to adaptation, a perceptual phenomenon typically observed for basic sensory properties such as color, contrast or speed [4]. Moreover, recent neuroimaging investigations have revealed that, in the adult brain, the extraction of purely numerical information from the visual scene is fast, automatic, and independent from the processing of other quantitative characteristics [5–7]. Taken together, these findings promote the existence of a genuine *number sense* in our species, whereby number is encoded as a primary property of the visual scene irrespective of its relevance and by means of dedicated processes.

Yet, number is inherently and inevitably correlated with other physical quantities such as size, density, or total surface. For example, a greater number of dots implies more surface occupied or, if the total surface is fixed, a smaller size of the dots. In behavioral assessments, despite the clever control conditions designed by various investigators (e.g. [8]), it remains difficult to disentangle the weight assigned to purely numerical information from that assigned to non-numerical dimensions[9–11]. Behavioral performance is modulated not only by the experimental task[12] but also, and crucially, by the age of the participants. For instance, the numerical judgments of younger children are highly susceptible to irrelevant quantitative features[10,13] and recent observations indicate that, provided balanced discriminability, surface area is more salient than number during childhood while the reverse pattern is typical of educated adults[14,15] . Therefore, despite multiple reports of approximate perception of number in infants[11,16] and neonates[17,18], the hypothesis that humans are born with a number sense remains debated, with authors either asking

for more evidence[19] or proposing that early numerical discrimination is best explained by a holistic, generalized and monolithic sense of magnitude [20–22].

The neurodevelopmental findings available to date remain insufficient to clarify this issue. In agreement with a crucial role for the intraparietal sulcus in adult numerical cognition[23], an EEG investigation on 3-month-olds[24] and two fNIRs studies on 6-month-olds[25,26] reported that numerical changes (relatively to shape changes) trigger specific activity over the right parietal lobe. The design of these experiments was based on unpredictable changes in the numerosity of visual displays while non-numerical dimensions were variable and poorly informative when taken across the entire stimulation set. However, such an approach has been criticized by the proponents of the generalized magnitude hypothesis[20]. In fact, a number change might only be detected when it correlates with some non-numerical dimension (e.g. total surface), eliciting a response that might be large enough not to be washed out by other trials in which that dimension is decorrelated from number[27,20]. Because in the parietal region neuronal populations involved in the processing non-numerical quantities (e.g. spatial length or object size) are spatially intermingled with number-specific ones[28–30], it is difficult to counter this argument. In human adults, neuronal populations that selectively encode number have been revealed using high-field fMRI [30,31] and intracranial electrodes[32] whereas, relative to the latter, the brain-imaging methods and analytic tools used in infants so far could offer only a much coarser precision.

The goal of the present study is to clarify the nature of numerical representations in human infancy. With this aim, we queried the existence of a genuine number sense in 3-month-olds. To address the challenges just described, we exploited the spatial and temporal richness of infant EEG[33,34] by combining an unusually dense coverage of the scalp (custom design net with 256 channels; Fig. S1) with multivariate pattern analyses (MVPA).

We based our design on two postulates. First, if the human brain regards number as a primary sensory descriptor, we would expect infants to encode numerical information separately from non-numerical parameters, just as adults [5–7]. That is, if the infant brain contains neuronal populations specifically tuned to number, a multivariate decoder trained on EEG signals should be able to discern which number the infant saw or heard, independently of the format of presentation. To test such prediction we coupled a strategical stimulus design with cross-condition decoding and assessed whether it is possible to retrieve the same neural code for number across physical variations. Complementarily, we used representation similarity analysis (RSA)[35] to separate the significant contributions of each physical parameter as a function

of time, thereby determining whether number can be effectively separated from the processing of concurrent non-numerical magnitudes.

Second, if the human brain regards number as a primary sensory descriptor, we would expect infants to extract numerical information automatically and pre-attentively, just as adults [5–7]. To the best of our knowledge such an eventuality has never been probed: in all previous developmental studies active attendance and artificially enhanced numerical saliency might have led to attentional biases precluding the possibility to characterize spontaneous encoding. To overcome such limitation, we evaluated number processing during sleep. Furthermore, unlike previous paradigms, our experimental design did not rely on change detection, nor elicit any implicit comparison emphasizing the number dimension. We presented auditory sequences varied in duration, tone rate, instrument and pitch; and visual displays heterogeneous for object size, density, identity and color. The stimuli were randomly intermixed, and number was only one of their many descriptive features. If the human brain regards numerosity as a primary perceptual descriptor since start, we expected very young infants to encode numerical information pre-attentively and irrespective of any concurrent non-numerical parameter.

RESULTS

Three-month-old infants (N=26) were exposed to 24 different types of auditory sequences composed of either 4 or 12 identical tones. The tones were chosen from eight possibilities (2 instruments x 4 notes) to create sensory variability outside the numerical and non-numerical parameters of interest. The sequences were played in a randomized order for a total of ~1900 trials/subject. Crucially, the auditory space (Fig. 1A) carefully controlled for all the quantitative dimensions involved (Table S1). The fundamental advantage offered by multivariate decoding consists in the possibility to test estimators on number contrasts (i.e. pairs of sequences) where tempos and durations are orthogonal to those used for their training, thereby canceling the problematic correlation between numerical and non-numerical parameters (Fig. 1C and S2).

Retrievability of neural codes for number when non-numerical parameters are canceled out

Within three distinct assays, algorithms were trained to separate one experimental condition from a composite class that included sequences matched in either rate (50% of cases) or duration. At test, the non-numerical quantities distinctive for one particular numerical class during training were prevented from yielding above-chance scores since they characterized either both numbers (A1, B1: 4- and 12-tone sequences have the same total duration; A2, C2: they have the same rate), the opposite number relatively

to the training set (B2, C1), or none of the test trials (this was the case of the fastest rate in schema B and the longest total duration in schema C). To avoid any confound with the ongoing auditory stimulation, the decoding analysis was applied at the end of the sequence, when the brain may compute total numerosity[36]. Specifically, for every schema in Fig. 1C we employed a series of classifiers trained on brief (10ms) consecutive windows following of the onset of the last sound. The results (Fig. 2) revealed significant decoding of number in a broad time window (~400-800ms) following sequence ending. Fig. 2A shows that all estimators trained and tested in between 440 and 750ms achieved above-chance scores, with the best classification performance observed at 610ms (N=26; M±SD=0.557±0.044, chance=0.5). Their generalization performance across time[37] yielded a square-shaped matrix, indicating an essentially stationary code (Fig. 2B). Further, classification dynamics were qualitatively similar across all tests (Fig. 2C and 3).

We also considered the non-numerical features, rate and duration, and trained new sets of estimators to separate pairs of conditions differing in tone rate but matched for sequence duration ("4L" vs "12M") and differing in total duration but matched for rate ("4M" vs "12M"). At test, the numerical distinctions characterizing the training sets were inverted ("12L" vs "4M" and "12S" vs "4L" respectively) to obtain classification scores attributable exclusively to rate or duration (Fig. 3). We observed no overlap between number and rate decoding (Fig. 3A), nor between number and duration decoding (Fig. 3B), demonstrating that successful performance in the main analysis (Fig. 2) was not driven by the fact that, on average, sequences of 4 tones were slower or shorter.

Number is encoded separately from other magnitudes

So far, we have demonstrated that "4" and "12" can be reliably discerned from infant neural responses during sleep, when the effects of specific non-numerical parameters are canceled out. This result highlights that the infant brain keeps track of numerical information in a pre-attentive, automatic manner. However, the nature of this mechanism remains somewhat unclear: does successful classification rely on a generalized code where numerical and non-numerical cues are fused[20,22]? Or are these dimensions factorized in neural signals? To evaluate the separability of the neural codes for number and non-numerical magnitude, we used RSA which enables to assess the effect of multiple dimensions at once [31]. At every time point from sequence offset onwards, we computed the correlational distance between the average responses evoked by each pair of auditory conditions. Next, using multiple regression, we modeled the resulting neural dissimilarity matrices as a linear combination of three theoretical matrices (Fig. 4A-top) depicting the dissimilarity of the sequences along their defining quantitative dimensions.

With this approach, we obtained three series of beta weights reflecting unique portions of the neural variance that each quantitative dimension explained *independently* from the other two. Fig. 4A (bottom) shows how rate, duration and number could be effectively disentangled. Crucially, number exerted the strongest degree of modulation over a relatively late time-window with a peak in beta weights observed at 600ms (N=26, 525-805ms: p_{clust}=0.0001; Fig. 4A). In full agreement with the time-course of the decoding performance (Fig. 2), this finding confirms the numerical nature of the neural patterns underlying classification by demonstrating that the infant brain estimates number *separately* from the other magnitudes, just as adults.

Additionally, number coefficients were significantly above zero over an earlier window (105-280ms: p_{clust}=0.0001), indicating the existence of a preliminary process not captured by the decoders. Considering that the auditory ERP was still unfolding during this period, a superimposition of neural sources pertaining to numerical information and other sensory parameters might explain the difference in sensitivity between decoding and RSA analyses. We hypothesized that, during the sequence itself and before reaching a representation of total number, the brain might constantly keep track of each additional sequence item. This could be implemented via the update of an analog accumulator [38], through a hippocampal "event code" [39], or by a matrix-based successor function in neuronal vector space [40]. Regardless of the exact mechanics involved, the occurrence of such a process implies that the gradually evolving number should be discernible from neural signals while successive tones are played. To test the latter eventuality, we focused on "12" trials and applied RSA to the neural signals following tones whose ordinal number was separated by a certain ratio. Supporting our interpretation, Fig 3B (bottom) shows that beta coefficients for number were significantly positive over an early time window (160-335ms; p_{clust}=0.0007) following the 3rd vs. 7th tone (ratio 1:2.33). We found a similar effect following the 4th vs 9th tones (ratio 1:2.25) but not the 3rd vs. 5th tones (ratio 1:1.67; Fig. S3). This range of successes and failures indicates that early numerical encoding is imprecise and shaped by Weber's law, corroborating previous behavioral observations in 6-month-olds [41,42].

Cross-decoding reveals a supra-modal neural code for number in the infant brain

Our results demonstrate that the infant brain treats number as a basic property of auditory sequences, not reducible to other non-numerical variables. Is this representation bounded to audition or is it more abstract? We examined whether the same numerical code extends across sensory modalities, presentation formats (sequential vs simultaneous), and vigilance/arousal state. We took advantage of the fact that some infants (N= 16) were awake and attentive before or after the sleeping auditory session.

During these periods, we presented them visual displays of 4 and 12 simultaneous objects (Fig. 1B). Object size, density, total visual occupation and luminance were controlled across images. We selected the decoders successful in isolating a number-specific neural pattern from the auditory trials (Fig.2), optimized their learning by training them iteratively on all possible auditory schemas (Fig. 1C/S2, see STAR Methods), and assessed their performance on visual responses (Fig. S4). Strikingly, estimators trained between 440 and 610ms from the last tone composing the auditory sequences performed reliably above chance from ~300 to 580ms relatively to image onset (Fig. 5, N=16: p_{clust}=0.005), with the best performance obtained at 370ms (e.g. for the classifier trained at 500ms: M±SD=0.588±0.076, chance=0.5). This result did not depend on the type of control used over non-numerical visual attributes, since similar scores were achieved on trials that equated extensive parameters (N=16; training 490-510ms, test 360-380ms: M±SD=0.565±0.092) or intensive parameters (mean M±SD=0.605±0.148; no significant difference: t=-0.853, p=0.41).

DISCUSSION

In the current study, we exposed preverbal infants to an auditory space composed of sequences of musical tones with balanced numerical and non-numerical quantitative parameters. We used multivariate analyses to isolate purely numerical processes from any modulatory effect ascribable to the other quantitative characteristics of the stimuli (single tone duration and inter-tone intervals, rate, sequence duration and total amount of sound/silence). Our results reveal that the infant brain encodes numerical information automatically and separately from other non-numerical dimensions, indicating that number is a fundamental and key dimension for representing the auditory environment. Such a finding is somewhat counterintuitive: after all, number encapsulates a discretization process, just as rate, and a cumulative aspect, just as duration. What could be the benefit of a primary neural mechanism specifically dedicated to the accumulation of discretized sensory evidence? The answer may lie in the unique representational flexibility number affords: unlike the other quantitative parameters, number can be abstracted away from sensory modalities, time and space [43]. Coherently with this conception, our cross-decoding analysis revealed that at ~12 weeks of age our brain engages a neural mechanism for number processing that transcends wakefulness state, input modality, and temporal versus spatial distribution.

While numerical behaviors have been well documented from the 6th month of age, less is known about younger infants. According to a recent habituation study, 4-month-olds are unable to detect a numerical change with ratio 1:3 unless they are provided with redundant multimodal stimuli [44]. Such an observation

favors the proposition that, if at all, young infants might encode numerical quantity only in particular circumstances of enhanced saliency [11]. In contrast, our analyses revealed numerical representations in the context of a rich and varied sensory stimulation (the sequences were composed of different musical notes played by two orchestral instruments) in which number discriminability, and thus saliency, was equated to that of the other quantitative dimensions (duration and rate). As with other dissociations between EEG and behavior in the literature [24], different outcomes might arise from two factors. First, our experiment benefits from a large number of trials and therefore optimal sensitivity relatively to the few measures obtainable with behavioral habituation paradigms. Second, while overt behaviors result from an intricate combination of multiple neural/cognitive processes and developmental immaturities, brain measures provide a more direct index of a given representational mechanism.

Our results do fit with two seminal reports pertaining to newborns. Namely, after a brief familiarization phase, 0- to 4-day-olds were found to preferentially look at visual arrays that matched an auditory sequence in number of items, provided that the numerical ratio between test displays was at least 1:3 [17,18]. Still, such a preference might have reflected an instinctive mapping between two arbitrary quantitative dimensions, such as rate and density, similar to those observed in both neonates [45] and older infants [46,47] and similar to other intuitive auditory-visual mappings connecting, for instance, shape spikiness to vowels[48]. If that was the case, detection of a correspondence between auditory temporally-distributed information and visual ensembles might arise from a generalized magnitude representation [20,22]. The neural evidence provided by the current study, however, supports the alternative interpretation according to which newborns can detect a genuine supramodal numerical correspondence; that is: they "perceive abstract numbers" [17]. Interestingly, these behavioral observations also suggest that the abstract neural code isolated by our analyses in 3-month-olds is likely to be operational since birth. Such a conclusion is corroborated by the computational demonstration that cells tuned to the numerosity of visual sets emerge automatically in untrained neural networks [40,49,50] and enable them to perform number discrimination tasks, even in the presence of misleading non-numerical quantitative parameters [50]. Such mechanism, extended to cross-modal stimuli, could explain the present findings. Furthermore, inspecting the internal dynamics of the neural networks uncovered the existence of number-sensitive and number-selective response profiles [49], whereby tuning emerges from monotonic increases and decreases of neuronal activity in the earlier layers of the network [50]. It is intriguing to notice how such "summation coding" is consistent with the accumulator captured by our RSA analysis (Fig. 4B).

Numerical representations in human infants appear similar to those reported for non-human animals. As mentioned in the introduction, the ability to rely on quantities is widespread across the animal kingdom, among bees[51], fishes[52], birds and mammals. In line with our observations neurons tuned to approximate number irrespective of other physical parameters, such as item size or density, have been discovered in the intraparietal sulcus of the macaque[3], in the nidopallium of naïve chicks [53] and in the pallium of zebrafish[54]. Within the intraparietal sulcus of the monkey, sequential numerosities are processed by two distinct neuronal populations. The first encodes intermediate numbers, with neurons progressively increasing their firing rates while stimuli presentation is ongoing. At the end of the enumeration process, the final cardinality is accomplished by a different neuronal population engaged by numerical information spread across both time and space[36]. These findings are consistent with both our retrieval of an early and late number effect (Fig. 4) and the biphasic dynamic of the neural networks mentioned above[49,50]. Lastly, the numerical representations of the rhesus macaque have been proven to be partially cross-modal[55,56]. Overall, these observations confirm that number constitutes a basic dimension that helps organisms structure the environment. We reported neural separability for 4 vs. 9 (Weber fraction=0.55) but not 3 vs 5 (Weber fraction=0.4) in congruency with the notion that the ratio between two discriminable quantities narrows down from 1:3 at birth [17,18] to 1:2 over the first semester[41,42] and continues to decrease with age. Similar Weber fractions (i.e. ~0.5) have been observed in crows[57] and monkeys[58], although using a different paradigm. Whereas for monkeys numerical acuity can be enhanced with training[59], that of humans might increase due to exposure, education, and brain maturation[60–62].

What about human adults? Paralleling our observations, steady-state visual evoked potentials disclosed the occurrence of spontaneous numerical encoding even for complex and heterogeneous stimuli that depart from the artificial-appearing dot arrays classically used in this research field[63]. Further, electrophysiological evidence indicates that the adult visual cortex might extract numerical information through a series of two processing stages, compatible with those highlighted by our RSA analysis [64]. Yet to our knowledge, an abstract neural code for non-symbolic, non-verbal numerosity as that isolated by our classifiers (i.e. a code that transcends format, modality and arousal state) has never been observed in humans. Two separate fMRI studies on adults have reported overlapping neural activations in response to visually and auditorily presented sequential numerical displays[65] and to sequentially and simultaneously presented visual numerical displays [66]. However, subsequent fMRI investigations failed to replicate these findings, suggesting a role for active comparison processes in the previous results [67,68]. Thus, the neuroimaging evidence alone leaves open the possibility that numerical representations become modality and format specific as we grow older. Nevertheless, adult psychophysics offer

meaningful insights in this regard. Adaptation of numerosity estimates [4] generalizes across formats and sensory modalities, from sequential streams to spatial arrays (and vice versa), and from auditory sequences to visual displays (again in a bidirectional fashion). Such cross-format and cross-modal adaptation effects are nearly as large as those observed within-format and within-modality, revealing that to be adapting is an abstract quantity system [69]. Further, adaptation remains spatially specific in all cases indicating a relatively basic encoding mechanism rather than a higher-level cognitive construct [70]. Overall, this pattern of behavioral findings provides a glimpse of a discrete number system in adults that parallels the one isolated here within a developmental neuroimaging setting.

In conclusion, our observations show the existence, early in life, of a spontaneous, specialized and modality-independent neural mechanism that encodes number as a basic sensory descriptor; that is to say: a *primitive and abstract number sense*. Given the widespread idea that approximate numerical computations function as a "start-up tool" for the acquisition of mathematics [71], our findings are likely to have practical repercussions in the context of educational and rehabilitative interventions [72].

ACKNOWLEDGMENTS: We thank Evelyn Eger for her advice on experimental design and François Leroy for hardware assistance. We are grateful to Marie Palu for setting up the initial phases of data collection. Finally, we are thankful to Ana Fló for guidance on preprocessing and Jean-Rémi King for providing coding tips during the analysis. We thank the parents, for their patience during the long recording sessions, and the infants, for their cooperation.

AUTHOR CONTRIBUTIONS: Conceptualization: GG, GDL ; Formal analysis: GG ; Investigation: GG, CV ; Methodology: GG ; Software: GG ; Supervision: GDL, SD ; Validation: GDL, SD ; Visualization: GG, GDL, SD ; Writing – original draft: GG, GDL, SD ; Writing – review & editing: GG, GDL, SD

DECLARATION OF INTERESTS: The authors declare no competing interests.

MAIN FIGURES CAPTIONS

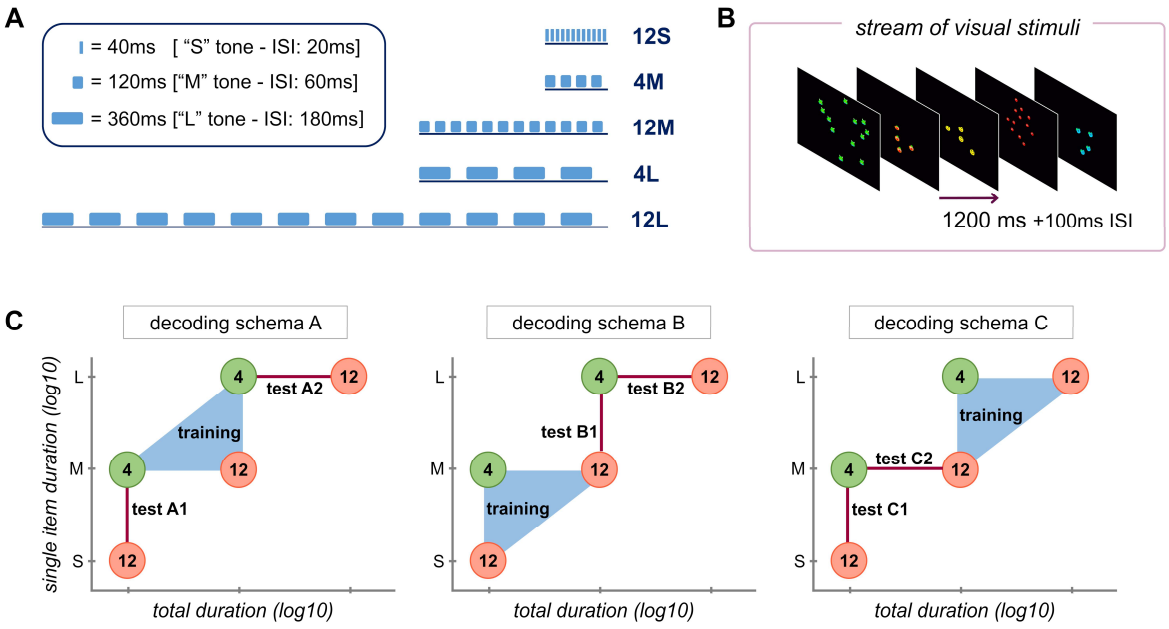

Fig. 1 Experimental paradigm for number decoding. (**A**) Sleeping 3-month-old infants were exposed to five types of auditory sequences containing 4 and 12 tones while recording 256-channel EEG. All quantitative parameters were controlled (see Table S1). (**B**) When feasible and in a separate session, infants were presented with images of 4 and 12 colorful objects. Object size, density, total occupation area and luminance were controlled (ISI=inter-stimulus-interval). (**C**) Scheme for training and testing number decoders. Each decoder was first trained to separate numerosities 4 versus 12 based on EEG data from three conditions (triangle), equalized for either total duration (x axis) or individual tone duration (i.e. rate, y axis). Each decoder was then tested for generalization on new trials, again matched for one or the other parameter. Tests B2 and C1 provide the strictest control for non-numerical quantities as they exclusively involve new conditions never seen during training; the latter are matched for one non-numerical parameter (tone duration for B2, total duration for C1) whereas the value on the other parameter would assign them to the wrong number class. For an alternative visualization of these schemas see Fig. S2.

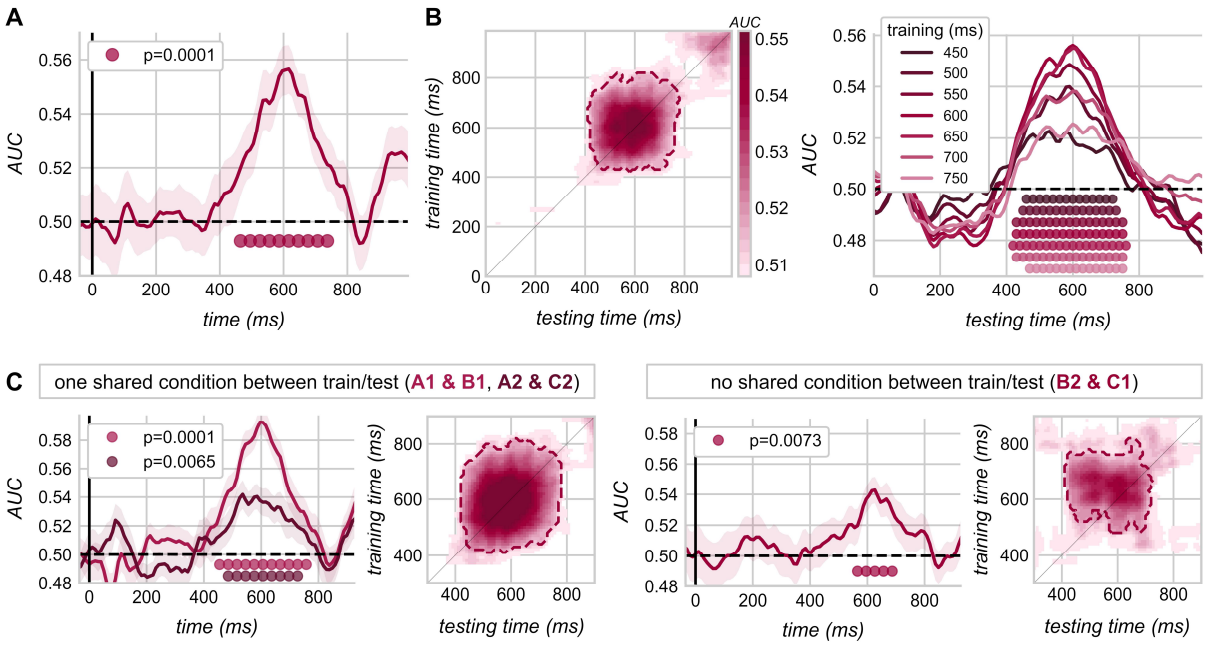

Fig. 2 Classification of "4" vs "12" from infant neural responses. (**A**) Average decoding performance of all classifiers (Fig. 1C). Time 0 corresponds to the onset of the last tone composing the sequence. Shaded areas indicate the standard error (SEM) across subjects (N=26) and dotted black lines mark theoretical chance level. (**B**) Generalization-across-time (GAT) matrix. Dashed contours delimit statistical significance, calculated by means of a cluster-based permutation t-test against chance (p_{clust}=0.0001). Right panel: slices through the GAT matrix show the performance of classifiers trained every 50ms between 450ms and 800ms. (**C**) Average performance of classifiers sorted as a function of whether they did or did not share any specific stimulation condition between training and testing (see Fig1C). Note that all classifiers were always cross-validated with a separate set of trials not used for training.

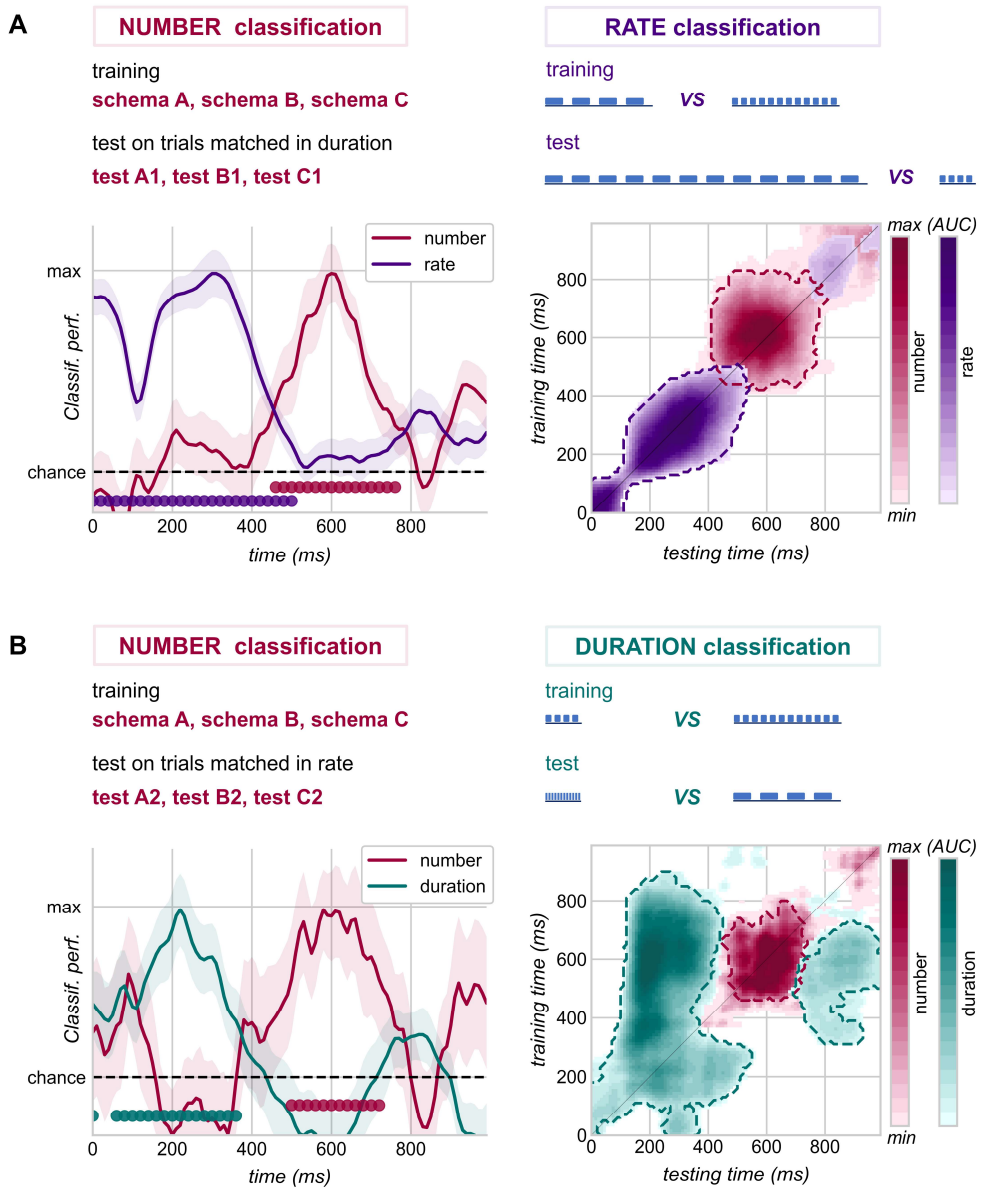

Fig. 3.

Classification of non-numerical parameters (rate and total sequence duration) relative to number (A) Within the training sets of our main decoding analysis, class "4" was necessarily characterized by auditory sequences that are *on average* slower relatively to those corresponding to class "12". With this qualitative assessment, we asked whether above-chance performance on conditions matched for duration (i.e. when duration-related effects are completely uninformative, see Fig. S2 for an intuitive illustration) might reflect the retrieval of a "slower than/faster than" type of computation. In other words, we evaluated whether successful number classification could derive from rate distinctions. In the two panels, rate-based classification performance (p_{clust}=0.0001) is superimposed over the average scores obtained with the main

decoding analysis when classifiers were tested on pairs of conditions that differed for both rate and number (p_{clust}=0.0002). We found no overlap between the two performances, indicating that the main classification scores were not contaminated by the retrieval of rate-related effects. (B) Within our main training sets, class "4" was necessarily characterized by auditory sequences that are *on average* shorter relatively to those corresponding to class "12". This qualitative assessment mirrors that of panel A. Namely, we tested whether successful performance on conditions matched for rate (i.e. when rate-related effects are completely uninformative) could reflect the fact that the infant brain encoded the stimuli in terms of "shorter than/longer than". As in A, the superimposition of classification scores reflecting pure duration-related effects (p_{clust}=0.0001 and p_{clust}=0.012) on the performance obtained in those main tests that included between-class duration discrepancies (p_{clust}=0.0123) indicates that the scores for number were not contaminated by duration-related processing.

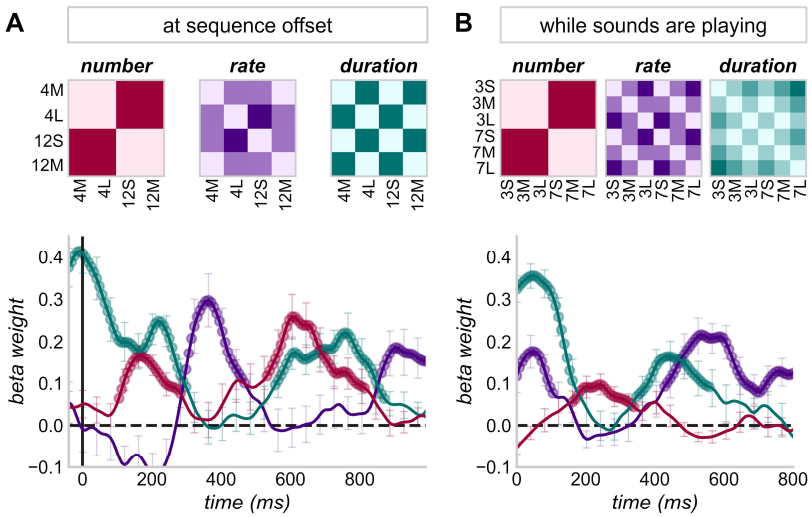

Fig. 4 Representation similarity analysis uncovers distinct dynamics for number and other quantitative dimensions. Theoretical matrices measuring the predicted distances between pairs of stimuli along the dimensions of number, duration and rate were entered in a multiple linear regression to explain the corresponding neural distances at each time point. Standardized beta weights averaged across subjects (N=26; vertical lines indicate the SEM) are marked by filled circles when significant. This analysis was performed either locked to the onset of the last tone of 4- and 12-item sequences (panel A), or locked to the onset of the 3rd and 7th tone within 12-item sequences (panel B). Numerical similarity predicts neural similarity in two distinct time windows (filled red circles): an early one (~200ms), present in both panels and reflecting current numerosity, and a late one (~600ms) present only at sequence ending and reflecting total numerosity.

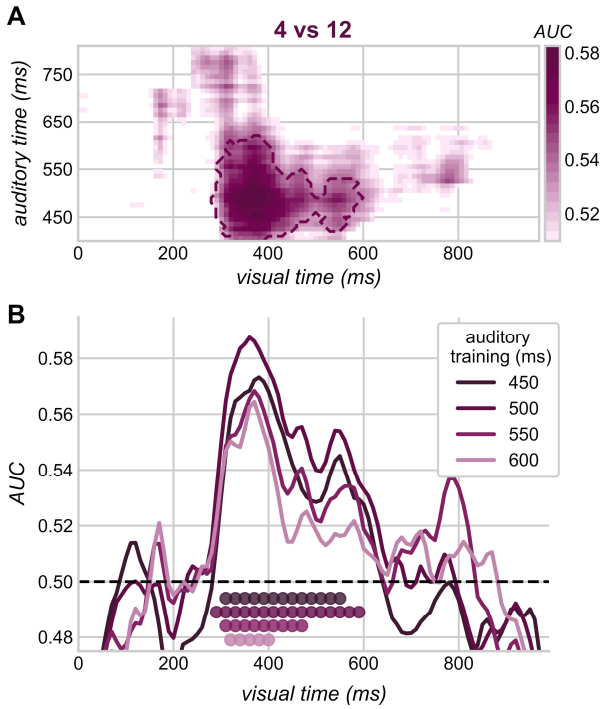

Fig. 5 Generalization of number classifiers from auditory to visual trials. (**A**) GAT matrix showing the performance of classifiers trained on auditory trials at different times relative to sequence offset (y-axis) and tested on visual trials at different times relative to visual onset (x-axis). Dashed contours delimit statistical significance (cluster-based permutation t-test). (**B**) Slices through the GAT matrix show the performance of classifiers trained every 50ms between 450ms and 600ms.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Giulia Gennari (giulia.gennari1991@gmail.com)

Materials availability

The study did not produce new materials.

Data and code availability

Original data and custom scripts are available upon request and will be deposited in a public repository according to the requirements of the journal.

SUBJECT DETAILS

26 healthy normal-hearing infants (15 females and 11 males) were tested between 12 and 14 weeks after birth (mean age= 13 weeks and 2 days). They were born after 37 weeks of gestation with a birth weight above the 10th percentile. An additional 23 participants were excluded from the analysis because of: impossibility to collect enough data due to excessive fussiness during the experimental session (n=13), insufficient number of trials after artifact rejection (n=4, the artifact rejection procedure is described below), equipment failure during data acquisition (n=6). The study comprised an auditory and a visual session. Out of the 26 infants included, only 16 (9 females and 7 males; mean age= 13 weeks and 2 days) had enough good-quality data for the subsidiary visual part. The protocol was approved by the regional ethical committee for biomedical research (CPP Region Centre Ouest 1). Parents gave their written informed consent before starting the experiment.

METHOD DETAILS

Stimuli

Auditory Stimuli: Orchestral string tones were synthesized with Ableton Live (software version 10; Berlin, Germany). Four notes (C_3, G_3, C_4 and G_4) were played by a viola and a cello with three possible durations (40, 120 and 360ms; referred to as "S" [short], "M" [medium] and "L" [long] respectively) for a total of 24 different sounds. The shape of the acoustic envelope was the same across all notes. Each auditory sequence comprised a single tone repeated 4 or 12 times at a constant rate. The duration of inter-tone intervals was half that of the single tone characterizing the sequence (i.e. 20ms for S sounds, 60ms for M sounds or 180ms for L sounds) resulting in three possible rates (1.9Hz, 5.6 Hz and 17Hz) and four possible total durations (240ms, 720ms, 2160ms and 6480ms; however the shortest total duration, i.e. "4S", was excluded a-priori from the analysis because it sounded awkward). "4" was chosen as the smallest number for the following reason. When it comes to encode numerical information, infants are thought to possess two systems: one for approximate quantities and another for tracking up to 3 objects in parallel[16,73].To date, the circumstances in which one or the other system are recruited to process small numerosities remain unclear (e.g. [41] vs [74]). In our study, we wanted to rule out the possibility that differences between numerical conditions could be attributed to the intervention of two distinct representational systems

rather than to genuine neural codes for number. Crucially, the minimal ratio between pairs of numbers, tone rates or total sequence durations was always 1:3. Given that infants differentiate numbers and durations with the same level of precision [75], this means that the three quantities were equally discriminable for our participants. All sequences had the same sound-to-silence ratio (2:1) and equal loudness (75 dB). They were recorded on the left audio channel, which was connected to the loudspeakers. A 10ms-square signal was positioned on the right channel in correspondence with the onset of the initial tone and with the offset of the final tone composing each sequence; these clicks were used as a TTL signal to ensure accurate synchronization between the EEG recording and the stimulation.

Visual Stimuli: The visual stimuli were the same as those used by Izard, Dehaene-Lambertz & Dehaene in 2008 [24]. They consisted of a set of 400 images depicting either 4 or 12 colorful animal-like objects (8 different objects) on a black background. To minimize any possible effect ascribable to non-numerical perceptual attributes, the position of the objects and the physical parameters of the image varied across stimuli following two rules. Namely, in 200 images the extensive parameters of the display (total luminance and total occupied area) were kept constant across numerosities, whereas in the remaining 200 images the intensive parameters (object surface size, average area devoted to each object) were equated between numerosities. For more details on the control for non-numerical factors see [24].

Procedure

Infants were tested in a soundproof Faraday cage equipped with a computer screen and loudspeakers on the ceiling. They were held by a caregiver in a comfortable position and constantly monitored by the experimenter from two video cameras located underneath and above the screen. All stimuli were presented using the Python package PsychoPy [76].

The study comprised long-lasting auditory sessions and a subsidiary visual stimulation. Whereas sleep was required during the former, participants needed to be awake and calm in order to attend the visual display. Therefore, the order of visual/auditory sessions was based on their alertness state when arriving at the lab: when the infant was alert the visual part was carried first and vice-versa in case of sleepiness. Of the 26 subjects included in the final analysis, 13 infants saw the images before the auditory stimulation and 6 after, 1 infant partly before and partly after the auditory sequences. The remaining 6 infants had no visual stimulation.

Auditory: Sequences were played with a fixed inter-sequence-interval of 1 second and organized in blocks of 688 trials (72 for “4S”, 136 for “4M”/ “4L”/ “12M” and 104 for “12S”/ “12L”) where the four notes, the two instruments and two numerosities were balanced. The different number of trials for each type of sequence was motivated by analytical plans (e.g. 4S was included in the stimulation to provide perceptual harmony but excluded from the analysis a-priori) and practical constraints (i.e. the need to collect a large number of trials in a reasonably limited amount of experimental time). The order of the sequences was randomized with two constraints: the same number could not be presented more than 4 times in a row and the same condition (characterized by a given note, instrument, numerosity, rate and duration) could not be repeated more than twice in a row. A minimum of 2 blocks (corresponding already to ~71 minutes of listening) and a maximum of 3 blocks were presented to each participant. Breaks were taken whenever necessary and sleep was strongly encouraged. Often, the auditory stimulation started while the infants were still awake. However, in case the participant did not fall asleep within a short time window (~8/10

minutes), the session was paused or terminated. On average, infants were asleep for 72% of the auditory session.

Visual: Images were presented for 1200ms in a continuous stream and randomized order, interspersed with 100ms-long blanks. They were organized in mini-blocks of 100 items, where numerosities and non-numerical parameter control strategies were balanced. The onset of each image was recorded through a photodiode capturing the appearance of a white rectangle at the bottom corner of the computer screen (not visible by the subject) and sent as a TTL signal to the EEG recording system (NetStation 5.3, EGI). When the infant looked away the stream of numerical displays was interrupted, and a colorful attractor was presented until attention was re-established. The visual stimulation ended after the presentation of all the 400 images available or as soon as the participant could no longer engage with the displays.

EEG recording

The electroencephalogram (EEG) was continuously digitized at 500 Hz (Net Amps 300 EGI amplifier combined with NetStation 5.3 software, EGI®, Eugene, OR, USA) from 256 channels. We used a prototype HydroCel net (EGI) referenced to the vertex. Twenty of the standard temporal channel locations in a classical geodesic 128-sensor net were replaced by tight grids of sensors (70 electrodes on each side, organized in hexagonal pods) with no sponge inserts (Fig. S1A). Electrodes were made of carbon fibers embedded within a plastic (ABS) substrate and coated with silver-chloride.

Data preprocessing

The data were first band-pass filtered ([0.5 - 40Hz]) and the mean voltage of each electrode was set to zero. We then followed an artifact detection-correction procedure validated and exhaustively described by Fló et al. [77]. Namely, we based artifact detection on adaptive (rather than absolute) thresholds to account for inter-individual variability and for the heterogeneous influence that distance to the reference and vigilance state exert on the voltage. Thresholds were set independently for each subject and for each electrode upon the distribution of different measures along the whole recording (threshold = median +/- n*IQ, where IQ is the interquartile range of the distribution). We used series of algorithms that rejected samples on the basis of: the voltage amplitude and its first derivative; the variance across a 500ms-long moving time window; the fast running average and the deviation between the fast and the slow running averages within a 500ms-long sliding time window. Two additional algorithms identified whether the power within the 0-10Hz band was excessively low or within 20-40Hz excessively high relative to the total power; and whether the voltage amplitude displayed by each sensor at a given time point was disproportionate relative to that recorded by the other sensors at the same instant. For these last two algorithms, thresholds were computed upon the distribution across channels.

Artifact detection was conducted on the continuous recording in an iterative fashion (4 loops in total). At each run, previously identified bad samples were kept aside for the subsequent artifact detection steps. We started by applying the algorithms twice in order to identify very short signal disruptions (80ms max), corresponding to heart beats or jumps. We corrected these very short segments by estimating their principal components (PCA) and removing the first *n* components determining 90% of the variance. This operation was followed by a high-pass filter (0.5Hz) to eliminate possible drifts that could have been created by this local correction. Next, we applied the detection algorithms twice more. At the end of these

iterations, we obtained a rejection matrix of the same size as the EEG recording indicating "bad" time-samples for each electrode.

The EEG data and the corresponding rejection matrix were then segmented into epochs from -40ms to +1240ms relative to the onset of the last tone composing the auditory sequences and from -200ms to +1300ms relative to the onset of the images. To analyze responses within sequence, a third set of epochs was created from the onset of the first note and included a different time-window depending on the sequence lengths: [-200 to 1820ms] for "4M" and "12S" trials; [-200 to 3260ms] for "4L" and "12M" trials; [-200 to 3260ms] and [+2500 to 6500ms] for "12L" trials. The longest sequences were divided in two parts, first from the onset of the first tone until the end of the 6th inter-tone interval and second from the onset of the 6th tone until the 12th inter-tone interval.

Once segmented, we used the rejection matrix to mark, in each epoch, the time points containing prominent artifacts (*bad times*) and channels that did not function properly (*bad channels*). Specifically, *bad times* were periods longer than 80ms with a percentage of rejected channels superior to 30% or beyond 2IQ from the 3rd quartile of the distribution of the percentage of rejected channels across time. Similarly, *bad channels* were the ones not working properly for more than 30% of the time points composing the epoch or with a percentage of bad samples that went beyond 2IQ from the 3rd quartile of the distribution of the percentage of rejected samples across channels. With this step we extended in time and space the detected bad samples based on the assumption that when a conspicuous portion of channels/times are detected as "bad", the temporal/spatial neighbors are likely to be affected by the artefact as well. *Bad channels* and long rejected segments of a given electrode were corrected using spherical splines interpolation [78] only if at least 50% of the neighboring channels were intact.
Epochs were discarded if more than 15% of their samples contained artifacts or if more than 2.5% of their channels were marked as bad. Since in infants ongoing activity is particularly large and can hinder the extraction of the evoked potential, epochs were also discarded based on their Euclidean distance from the average, i.e. when their mean or maximum distance from the average response was an outlier in the distribution (> 3rdquartile + 1.5*IQ). Following these automated rejection steps, the remaining epochs were visually inspected and a few channels or trials still presenting obvious aberrancies were dropped. Accepted epochs were low-pass filtered at 20Hz and mathematically re-referenced to the mean of all channels.

Since our paradigm was mainly based on the auditory stimulation, our inclusion criterion concerned the latter: participants were included in the study with a minimum of 192 artifact-free epochs for *each* of the most frequent auditory conditions ("4M", "4L", "12M"). This criterion was set to guarantee a minimum of 24 pseudo-trials for every class in each training set (all the details are explained in the *Decoding* section below). In our final group of infants (N=26), the mean rejection rate for auditory trials was 33.5% (with median=34.3% and range 12.4 to 48.4%). The mean rejection rate during wakefulness was 59.6% (with a median of 67.1% and range 28.2 to 80.8%). Considering that participants were awake during less than 30% of the auditory session, this implies that during the vast majority of trials included in the main analyses infants were asleep. On average, the number of artifact-free epochs available per subject was 1252, including 247 trials for "4M"/"4L"/"12M" and 189 trials for "12S"/"12L" (the remaining epochs belonged to the "4S" condition and were discarded from further consideration).

The visual part of the experiment was much shorter due to the limited attention span typical of this age. For an infant to be included in the cross-modal analysis, we required at least 64 artifact-free epochs. The mean rejection rate for visual epochs was 62% (25 to 77.4%) and only 16 out the 20 subjects who attended the visual displays met the criterion. The average number of artifact-free visual epochs available for these 16 subjects was 109.

QUANTIFICATION AND STATISTICAL ANALYSIS

Decoding

Time-resolved multivariate pattern analyses were conducted within subject, relying on the Python packages MNE [79,80] and Scikit-Learn [81]. EEG data were first prepared by removing a linear trend from the entire segments to remove eventual slow drifts. After this preliminary step, epochs were divided into 110 consecutive windows of 10ms, from -40ms to 1060ms relative to the onset of the last tone composing the auditory sequence. All the procedures described in this section were carried at the level of single time-windows, each corresponding to a matrix with shape *n* channels *x* 5 samples (sampling rate=500Hz, 5 samples=10ms). The general goal of the decoding analyses was to predict a vector of binary categorical data (*y,* containing the classes "4" vs "12") from a matrix of single-trial neural data (*X)* which included all EEG channels.

For the main analysis we used three separate sets of estimators and followed the three complementary strategies illustrated by Fig. 1C and S2. In each training phase, one number class included trials belonging to a single experimental condition (e.g. "12M"), while the alternative numerical class was composed of two experimental conditions, one characterized by the same tone rate (e.g. "4M") and the other characterized by the same total duration (e.g. "4L") present in the homogeneous class (in this example: "12"). With this design, we minimized the impact of non-numerical parameters on the training process as number (4 vs. 12) was the only reliable feature to separate classes. Further, to make sure that duration-based or rate-based learning could not lead to successful performance, each set of classifiers was tested twice, on two different datasets (Fig. 1C and S2). In a first test (A1, B1 & C1) all sequences had the same total duration while the specific rate indicative of number during training could not lead to above-chance scores since it was either not at all present (A1, B1) or a peculiarity of the opposite numerosity (thus misleading, C1). Specifically, test A1 used short tones for 12 (fast rate) and medium tones for 4 (medium rate) whereas during training the fast rate was absent and the medium rate corresponded to both 4 and 12 (i.e., during training, the medium rate was uninformative to class separation). Test B1 entailed a similar configuration. Concerning C1, the medium rate characterized "4" while it corresponded to "12" during training. The second test (A2, B2 & C2 in Fig. 1C and S2) followed the same logic, in a reversed fashion: tone rate could not drive above chance scores since it was the same across conditions and the total duration indicative of the composite class during training could not lead to above-chance scores since it was either absent (A2, C2) or a peculiarity of the opposite numerosity (thus misleading, B2). Note how sub-schemas B2 and C1 yield pure cross-condition performances as both test conditions correspond to new sequence types, never employed during training.

In order to avoid overfitting, we used a cross-validation procedure with 100 loops. At each run, trials were shuffled, then assigned to the respective training and test sets. To ensure equal contribution of each experimental condition and at the same time maximize the number of trials during training, the splitting

was always organized to ensure balanced composite classes (in terms of *n* epochs per condition) but still exploit all 4M/4L/12M trials available for a given subject. Concerning schema A, the number of "4M" trials was first equated to that of "4L" (by randomly selecting and dropping *n* extra epochs for the most numerous condition), then 15% of "4M" trials and 15% of "4L" trials were kept aside for the test phase. The splitting was slightly different in schemas B and C in order to counterbalance the fact that "12S"/"12L" trials were less numerous relative to "12M" (see *Procedure* above). Namely, 80% and 20% of "12S"/"12L" trials were assigned to the training and test set respectively. The splitting of "12M" trials was then calibrated to obtain a balanced training set in terms of number of epochs per "12" condition (e.g. for schema B: *n* 12M trials in test set = total number of 12M trials available – *n* of "12L" trials in training set). Partitioning was always performed in a stratified fashion such that all sources of irrelevant variability (i.e. musical notes and instruments) were distributed in equal proportions. When a specific condition was used only within training or exclusively at test, all the corresponding trials were assigned to one of the two sets according to the schema at hand.

Once established the training and the test set for a given run, we applied a "micro-averaging" procedure, a strategy commonly employed to improve signal-to-noise ratio [82,83]. Within each experimental condition, this consisted in shuffling the epochs and then forming pseudo-trials by averaging together (randomly-defined) groups of 8 trials. At the end of such operation, we balanced the test sets by equalizing the number of micro-averaged epochs across numerosity classes. In practice, we randomly selected the same amount of pseudo-trials available for the least numerous class from the most abundant.

Next, following the z-scoring of each channel and time point across trials, we fitted a L2-norm regularized Logistic Regression to the training set [84] in order to find the hyperplane that could maximally predict *y* from *X* while minimizing a loss function. Since composite classes contained more trials than heterogeneous ones, a weighting procedure was applied in order to equalize the contribution of each class to the definition of the hyperplane. The other model parameters were kept to their default values as provided by the Scikit-learn package.

After training, the models were used to predict *y* from the test set and their performance was evaluated by comparing estimates to the ground truth. All algorithms produced, as an outcome, vectors of probabilistic estimates. These probabilities were scored by computing the area under the Receiver Operating Characteristic curve (AUC), which summarizes the ratio between true and false positives. The value of AUC ranges between 0 and 1, with 0.5 corresponding to chance level. The scores obtained across loops and from either all (Fig. 2A-B) or a group of train/test schemas (Fig. 2C) were averaged within subject before submitting them to statistical analysis (see *Statistical Analysis* below).

Generalization across time (GAT): Within the same cross-validation, estimators were tested both at the trained time sample and on all the other 109 windows. The outcome of this procedure is a temporal generalization matrix [37] where each row reports the classification scores of a single estimator trained at time *t* and tested all along the time-samples of the trial (each time lag *t'* corresponds to one column). When a neural code is sustained or recursive, a successful estimator trained at a given time point (i.e. specific to a given pattern of brain activity) achieves above-chance scores not only at the same time point but also at other time lags. Thus, the shape of the generalization performance within the temporal matrix can provide rich insights upon the dynamics of the neural activity patterns enabling classification.

Generalization across sensory modalities

In a second decoding analysis, we investigated whether the infant brain processes the numerosity embedded in auditory and visual displays through a common neural code. Given the limited attentional span of 3-month-olds and the peculiarities of our paradigm (i.e. relatively long image duration) it was impossible, in the current experiment, to collect enough visual trials to build robust estimators. Thus, training was always performed on auditory data. We used the same pipeline as above (100 cross-validation loops, L2-regularized logistic regression with weighted class contribution etc.) but this time probed decoder's ability to predict *y* from the neural responses to the visual displays. Given the divergence of training and test data, this analysis entailed the opportunity to employ all auditory conditions at once, with the potential benefit of increasing predictive power. Yet, in order to prevent class separation based on non-numerical parameters, it remained crucial to keep our three training schemas (Fig. 1C and S2) separate. Following such considerations, in order to maximize both sensitivity and specificity, we exploited an inherent property of the learning process: iterations. That is, the optimal hyperplane is computed though successive, intermediate and approximate minimizations of the cost function, while the model is updated incrementally after each pass over the dataset. In a standard decoding pipeline (initiated by the method model.fit() in scikit-learn), such incremental updates occur under the hood and classifiers are always 'fed' with the entire training dataset.

Building on the iterative nature of the learning process, we used a single set of decoders (one estimator for each of the time lags that led to successful classification in the main analysis) and trained them in an online fashion. Specifically, the initial pipeline was modified such that the training set (i.e. schema) changed (randomly) at each internal iteration, for a total of 600 *partial fits* (model.partial_fit() in scikit-learn). In practice, each classifier was 'fed' with training set A 200 times, training set B 200 times and training set C 200 times, in random order and while retaining the intermediate coefficient at each loop. The final weights of the model (those submitted to testing) corresponded to the average value of the coefficients computed across all updates. Overall, this strategy enabled us to capitalize on the possibility to use a larger training set while still minimizing the impact of non-numerical parameters on learning (by adopting the same training logic as the main analysis).

Visual data was prepared for tests in the same manner as the auditory data. Before micro-averaging, we equalized the amount of trials controlled for extensive parameters to that of trials controlled for intensive properties within each numerosity condition. When the trials available were too scarce to obtain a minimum of 5 pseudo-trials/visual numerosity (4 subjects), we implemented the micro-averaging using some of the single epochs more than once, with the constraint that two pseudo-trials could not share more than 2 single epochs (out of 8).

Each classifier trained on a given time-window *t* (in between 400 and 800ms after the onset of the last tones, i.e. within the period supporting number decoding for the auditory modality) was tested at every time-lag from 0 to 1000ms after the onset of the image (x-axis in Fig. 5A). Obtaining such temporal generalization matrix was essential for this analysis since we had no a-priori hypothesis concerning the temporal delay of numerical estimation within the visual modality.

Finally, to exclude the eventuality of non-numerical confounds on the observed performance, we created two supplementary test sets: one exclusively composed of trials with extensive parameter control (in

which object size and area devoted to each object co-varied with number) and the other including only those trials controlled for intensive elements (where total occupied area and luminance increased as a function of numerosity). Focusing on peak scores (Fig. 5) and once ascertained their normal distribution, we compared the performances attained on these two separate sets by means of paired sample t-tests (two-sided). In the main text we report the outcome obtained by averaging the scores over a 30ms squared window (3 *training time points* x 3 *testing time points*). With alternative solutions we observed similar null results. Namely, t-tests were also performed over a single performance (covering a time lag of 10ms): training at 500ms, test at 370ms, N=16 - mean AUC with fixed extensive parameters=0. 563±0.097, mean AUC with fixed intensive parameters=0.605±0.153, t=-0.832, p=0.42; and over a wider window (5 *training times* x 5 *testing times* i.e. 50ms): training 480-520ms, test 350-390ms - mean AUC with fixed extensive parameters=0.56±0.088, mean AUC with fixed intensive parameters=0.602±0.144, t=-0.89, p=0.39.

Representation Similarity Analysis

The aim of this analysis was to test whether numerical and non-numerical information could be dissociated from the activity patterns evoked by the auditory sequences. Crucially, unlike classification-based decoding, Representational Similarity Analysis (RSA) allows to assess the effect of multiple quantitative variables at once [31,35]. The general outline of the analysis consisted in modelling a set of neural (i.e. empirical) dissimilarity matrices, one for each time point, as a linear combination of 3 theoretical matrices providing all together an exhaustive description of the quantitative information embedded in the auditory space.

To compute neural dissimilarity, we started by down-sampling the EEG recordings (with a moving average of 4 time points) to 125Hz, then averaged together the epochs belonging to the same condition. Given that the potential of this kind of analysis is best expressed with rich experimental designs [35], we averaged trials where notes were played by different instruments separately. That is, for each main condition (Fig. 1A) we obtained two evoked responses, corresponding to the sub-conditions “viola” and “cello”. Finally, we calculated the correlational distance (1-Pearson across channels) between each pair of sub-condition. To counterbalance the fact that “12S/12L” trials were less numerous, we repeated this computation 100 times. At each loop, the evoked responses were calculated by averaging an equal number of trials per sub-condition (for each of sub-condition, we randomly selected the same number of trials available for the least abundant one). In this way, we made sure that each condition had the same signal-to-noise ratio and still exploited all the data available for a given subject, thereby optimizing the stability of the estimates. The final neural dissimilarities corresponded to the mean distances obtained across the 100 loops.

The theoretical dissimilarity matrices encoded the distance, on a logarithmic scale, between each pair of sub-conditions along the quantitative dimensions defining the auditory sequences: number, tone rate and total sequence duration. The three matrices were entered as predictors in a linear multiple regression in order to explain the neural distances observed at each time point. All the dissimilarity matrices were z-scored before estimating the regression coefficients. As a final outcome and for every subject, we obtained a set of beta weights reflecting the portion of the variance that each of predictor matrices *uniquely* explained in the evoked activity patterns over time. At the group level, significantly above-zero

beta weights imply that, over a certain window within the trial, one quantitative dimension modulated neural activity beyond the other two (i.e., when the effect of the other two was accounted for).

We performed two RSA analyses: one *at sequence offset* (i.e. from the onset of the last note), mirroring the decoding analysis, and one *within sequence*, to query the existence of a continuous numerical process that updates at each new tone.

Concerning the RSA performed at sequence offset (Fig. 4A), "12L" trials were excluded from the analysis in order to balance the design. Thanks to such expedient, all three predictors remained adequately decorrelated (number-rate: 0.26, number-duration: -0.17, rate-duration: 0.26) and their variance inflation factors (VIF) satisfactorily low (1.143, 1.19, 1.143 for number, tone rate and total sequence duration respectively). Note that it is important to minimize multicollinearity, as high correlations between predictors/high VIFs can compromise the reliability of the outcome coefficients.

For the RSAs performed within sequence (Fig. 4B and S3), the analysis was restricted to "12" trials and no exclusion was needed in order to obtain balanced distance matrices. Starting from the epochs crafted around sequence onset (see *Data preprocessing*), we obtained two sets of evoked activity patterns, corresponding to two cardinalities, by cropping the signal from the onset of the n^{th} note and up to 800ms thereafter. For the main analysis we chose to contrast "3" and "7" in order to parallel the RSA at sequence offset in the best way possible, thereby obtaining a set of beta weights that could be interpreted in relation to the former. For this contrast, total sequence duration (in this case: the time elapsed from sequence onset up to the specific cardinality) was equal between "7S" and "3M" (i.e. 360ms) and between "7M" and "3L" (i.e. 1080ms), mirroring the correspondence in the total durations of "12S"/"4M" and "12M"/"4L" (Fig. 1A). As in the previous case, this characteristic contributed to keeping multicollinearity at the minimum (VIFs were 1.059, 1.565, 1.555 for number, rate and duration respectively). Further, selecting the cardinalities "3" and "7" enabled to (a) minimize low-level effects related to sequence onset (e.g. habituation), (b) focus on a portion of the signal that was sufficiently far from sequence offset not to overlap with the main RSA analysis (c) test a numerical ratio greater than 1:2, taking into account that, according to behavioral observations, newborns require a ration greater than 1:2 to discriminate numerical displays [17]. Overall, the analytical choice of "3" vs "7" granted conceptual comparability between the outcomes illustrated by Fig. 4A and 4B.

Statistical analysis

We performed second-level tests across subjects employing the MNE dedicated functions. Following a standard approach in adult studies (e.g. [85]), we used one-sample cluster-based permutation t-tests [86] which intrinsically account for multiple comparisons over time (as a reminder: all our analyses were performed on consecutive time slices all along the trial corresponding to 10ms for decoding and 4ms in the RSA) . We tested whether (a) time-resolved classification scores were higher than chance and (b) whether multiple regression beta-weights differed from zero. The analyses considered two-dimensional clusters for decoding scores (i.e. they were always performed on the entire temporal generalization matrix) and one-dimensional clusters in the case of regression coefficients. Univariate t-values were calculated for every score/beta-weight with the exclusion of those corresponding to the baseline period. All samples exceeding the 95th quantile were then grouped into clusters based on temporal adjacency. Cluster-level test statistics corresponded to the sum of t-values within each cluster. Their significance was

computed by means of the Monte-Carlo method: they were compared to a null distribution of test statistics created by drawing 10 000 random sign flips of the observed outcomes. A cluster was considered as significant when its p-value was below 0.05.

SUPPLEMENTAL FIGURES AND TABLE

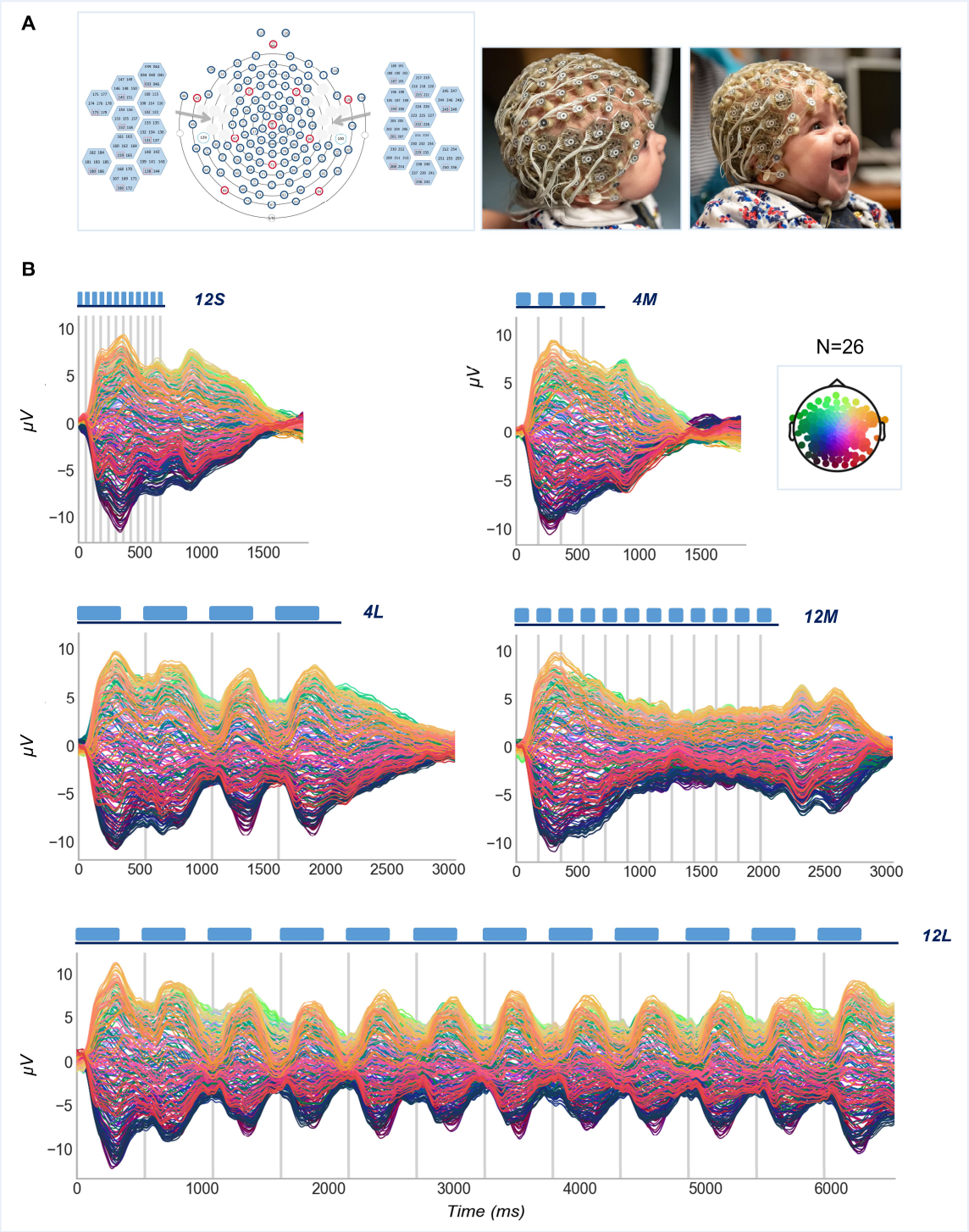

Figure S1. (A) Prototype EEG net featuring an intensive coverage of temporal areas; pictures adapted with permission from Gennari et al. [83]. (B) Grand Average high-resolution event-related potentials (ERPs; N=26) for the five types of auditory sequences. Vertical gray lines correspond to tone-onsets.

Figure S2. Tactical combination of training and test sets. Alternative visualization of the decoding strategy illustrated in Fig. 1C.

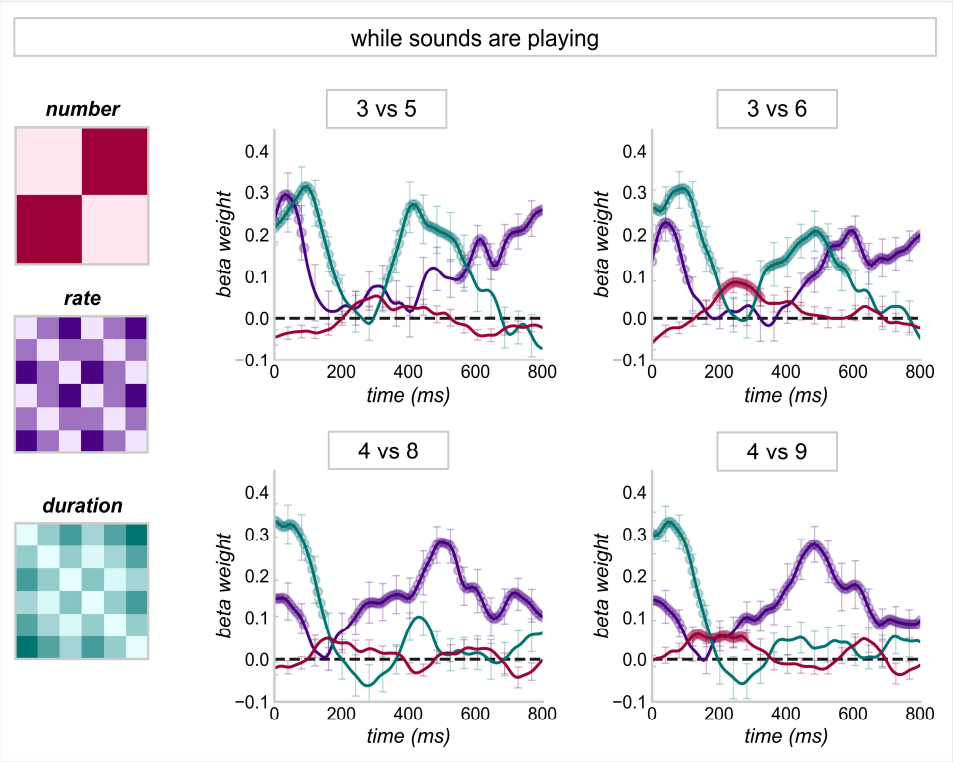

Figure S3. RSA within 12-tones sequences indicates that online numerical accumulation is imprecise. Multiple regression analyses mirroring that in Fig. 4B in all aspects except for the numerical contrast under investigation. Standardized beta weights are averaged across subjects (N=26; vertical lines indicate the SEM) and marked by filled circles when significantly above zero ("3" vs "4" – rate: 0-90ms pclust=0.0001 and from 545ms onwards p_{clust}= 0.01; duration: 0-185 and 335-560ms, p_{clust}=0.0001 | "3" vs "6" – number: 190-320ms p_{clust}=0.0054; rate: 0-105ms and from 455ms onwards, p_{clust}< 0.01; duration: 0-185 and 335-575ms, p_{clust}=0.0001 | "4" vs "8" – rate: 0-90ms p_{clust}=0.0001 and from 230ms onwards p_{clust}= 0.0035; duration: 0-145 p_{clust}=0.0005 | "4" vs "9" – number: 110-270ms p_{clust}=0.007; rate: 0-90ms and from 256ms onwards, p_{clust}< 0.01; duration: 0-160 p_{clust}=0.0001). In addition to these contrasts, the same RSA analysis was performed with "3" vs "9" (same numerical ratio as 4 and 12). The latter yielded a very similar beta-weights pattern to that observed for "3" vs "7" (Fig. 4B) and "3" vs "6" i.e. – number: 160-350ms p_{clust}=0.0006; rate: 440-615ms p_{clust}= 0.008; duration: 0-185 and 370-640ms, p_{clust}<0.001.

Note that these analyses were performed on data that were not independent (i.e. trials corresponding to the same sequences were aligned on different onsets). Regardless, "3" vs "6" and "4" vs "9" yielded similar results, with a significant time-window for the number regressor at around 200ms. Importantly, such modulatory effect is observed for "4" vs "9" but not "3" vs "5" revealing that the main result in Fig. 4B is not attributable to the intervention of an object-tracking system (for 3 but not 7 [16]) and that the accumulator mechanism is imprecise (i.e. a ratio <1:2 is insufficient to discern a numerosity effect).

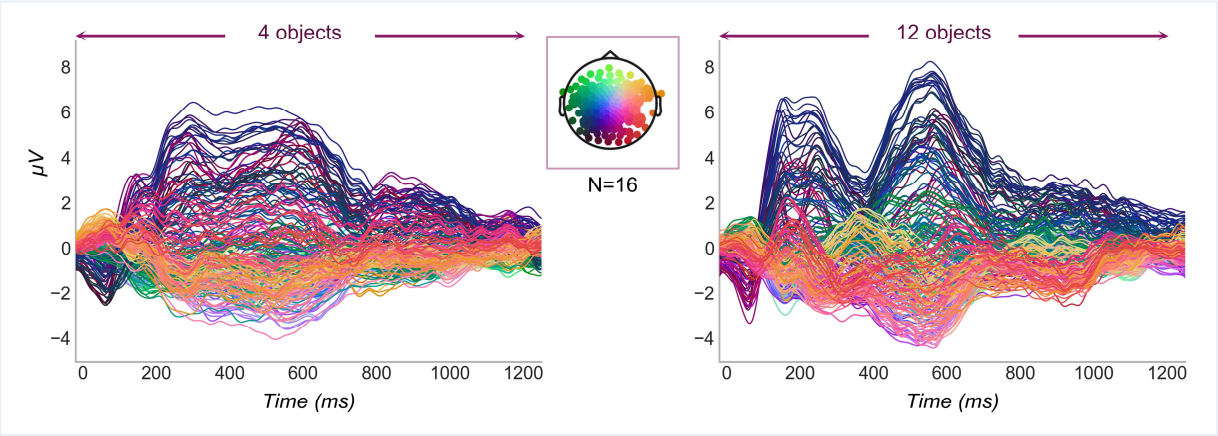

Figure S4. Grand Average high-resolution ERPs for the visual displays, divided by number condition. N=number of infant participants.

Condition / Parameters	4M	4L	12M	12S	12L
tone duration (ms)	120	360	120	40	360
inter-tone-interval (ms)	60	180	60	20	180
total sequence duration (ms)	720	2160	2160	720	6480
tone rate (Hz)	5.6	1.9	5.6	17	1.9
total amount of sound (ms)	480	1440	1440	480	4320
total amount of silence (ms)	240	720	720	240	2160

Table S1: Physical parameters characterizing each auditory condition. The auditory space was constructed in a way that ensured full and straightforward traceability of all the quantitative dimensions involved. Namely, a given tone duration is always matched to a specific inter-tone-interval and their ratio is constant (2:1), leading to a one-to-one correspondence between single tone duration and rate. As a consequence, when sequences of 4 vs 12 sounds were matched for their total duration (e.g. 12M and 4L), they also embedded the same amount of sound (e.g. 1440ms) and silence (e.g. 720ms). This design was crucial because it ensured the possibility to separate numerosity from any other possible quantitative parameter characterizing the stimulation. Any time our analysis controls for tone rate, single tone duration and inter-tone-interval are also accounted for and any time our analysis accounts for sequence duration, the total amount of sound and the total amount of silence are also accounted for. This design does not enable to disentangle the effects of single tone duration from that of inter-tone-interval and the effects of the overall sequence duration from those of the total amount of sound/silence: this kind of investigation was beyond the scope of the study.

REFERENCES

1. Spelke, E.S., and Kinzler, K.D. (2007). Core knowledge. Dev. Sci. *10*, 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x.

2. Dehaene, S. (2011). The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition (Oxford University Press, USA).

3. Nieder, A. (2021). Neuroethology of number sense across the animal kingdom. J. Exp. Biol. *224*, jeb218289. 10.1242/jeb.218289.

4. Burr, D.C., and Ross, J. (2008). A Visual Sense of Number. Curr. Biol. *18*, 425–428. 10.1016/j.cub.2008.02.052.

5. Park, J., DeWind, N.K., Woldorff, M.G., and Brannon, E.M. (2016). Rapid and Direct Encoding of Numerosity in the Visual Stream. Cereb. Cortex *26*, 748–763. 10.1093/cercor/bhv017.

6. Van Rinsveld, A., Guillaume, M., Kohler, P.J., Schiltz, C., Gevers, W., and Content, A. (2020). The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged EEG. Proc. Natl. Acad. Sci. *117*, 5726–5732. 10.1073/pnas.1917849117.

7. Lucero, C., Brookshire, G., Sava-Segal, C., Bottini, R., Goldin-Meadow, S., Vogel, E.K., and Casasanto, D. (2020). Unconscious Number Discrimination in the Human Visual System. Cereb. Cortex *30*, 5821–5829. 10.1093/cercor/bhaa155.

8. Xu, F., and Spelke, E.S. (2000). Large number discrimination in 6-month-old infants. Cognition *74*, B1–B11. 10.1016/S0010-0277(99)00066-9.

9. Mix, K.S., Huttenlocher, J., and Levine, S.C. (2002). Multiple cues for quantification in infancy: Is number one of them? Psychol. Bull. *128*, 278–294. 10.1037/0033-2909.128.2.278.

10. Rousselle, L., Palmers, E., and Noël, M.-P. (2004). Magnitude comparison in preschoolers: what counts? Influence of perceptual variables. J. Exp. Child Psychol. *87*, 57–84. 10.1016/j.jecp.2003.10.005.

11. Cantrell, L., and Smith, L.B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. Cognition *128*, 331–352. 10.1016/j.cognition.2013.04.008.

12. Leibovich-Raveh, T., Stein, I., Henik, A., and Salti, M. (2018). Number and Continuous Magnitude Processing Depends on Task Goals and Numerosity Ratio. J. Cogn. *1*, 19. 10.5334/joc.22.

13. Soltész, F., Szűcs, D., and Szűcs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. Behav. Brain Funct. *6*, 13. 10.1186/1744-9081-6-13.

14. Aulet, L.S., and Lourenco, S.F. (2021). The relative salience of numerical and non-numerical dimensions shifts over development: A re-analysis of Tomlinson, DeWind, and Brannon (2020). Cognition *210*, 104610. 10.1016/j.cognition.2021.104610.

15. Aulet, L.S., and Lourenco, S.F. (2021). No intrinsic number bias: evaluating the role of perceptual discriminability in magnitude categorization (PsyArXiv) 10.31234/osf.io/eh5pb.

16. Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. Trends Cogn. Sci. *8*, 307–314. 10.1016/j.tics.2004.05.002.

17. Izard, V., Sann, C., Spelke, E.S., and Streri, A. (2009). Newborn infants perceive abstract numbers. Proc. Natl. Acad. Sci. *106*, 10382–10385. 10.1073/pnas.0812142106.

18. Coubart, A., Izard, V., Spelke, E.S., Marie, J., and Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. Dev. Sci. *17*, 11–22. 10.1111/desc.12108.

19. Smyth, R.E., and Ansari, D. (2020). Do infants have a sense of numerosity? A p-curve analysis of infant numerosity discrimination studies. Dev. Sci. *23*, e12897. https://doi.org/10.1111/desc.12897.

20. Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2017). From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. Behav. Brain Sci. *40*. 10.1017/S0140525X16000960.

21. Hamamouche, K., and Cordes, S. (2019). Number, time, and space are not singularly represented: Evidence against a common magnitude system beyond early childhood. Psychon. Bull. Rev. *26*, 833–854. 10.3758/s13423-018-1561-3.

22. Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. Trends Cogn. Sci. *7*, 483–488. 10.1016/j.tics.2003.09.002.

23. Cantlon, J.F., Platt, M.L., and Brannon, E.M. (2009). Beyond the number domain. Trends Cogn. Sci. *13*, 83–91. 10.1016/j.tics.2008.11.007.

24. Izard, V., Dehaene-Lambertz, G., and Dehaene, S. (2008). Distinct Cerebral Pathways for Object Identity and Number in Human Infants. PLOS Biol. *6*, e11. 10.1371/journal.pbio.0060011.

25. Hyde, D.C., Boas, D.A., Blair, C., and Carey, S. (2010). Near-infrared spectroscopy shows right parietal specialization for number in pre-verbal infants. NeuroImage *53*, 647–652. 10.1016/j.neuroimage.2010.06.030.

26. Edwards, L.A., Wagner, J.B., Simon, C.E., and Hyde, D.C. (2016). Functional brain organization for number processing in pre-verbal infants. Dev. Sci. *19*, 757–769. 10.1111/desc.12333.

27. Gebuis, T., Cohen Kadosh, R., and Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. Acta Psychol. (Amst.) *171*, 17–35. 10.1016/j.actpsy.2016.09.003.

28. Sokolowski, H.M., Fias, W., Bosah Ononye, C., and Ansari, D. (2017). Are numbers grounded in a general magnitude processing system? A functional neuroimaging meta-analysis. Neuropsychologia *105*, 50–69. 10.1016/j.neuropsychologia.2017.01.019.

29. Borghesani, V., de Hevia, M.D., Viarouge, A., Pinheiro-Chagas, P., Eger, E., and Piazza, M. (2019). Processing number and length in the parietal cortex: Sharing resources, not a common code. Cortex *114*, 17–27. 10.1016/j.cortex.2018.07.017.

30. Harvey, B.M., Fracasso, A., Petridou, N., and Dumoulin, S.O. (2015). Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. Proc. Natl. Acad. Sci. *112*, 13525–13530. 10.1073/pnas.1515414112.

31. Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., and Eger, E. (2019). Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. eLife *8*, e45160. 10.7554/eLife.45160.

32. Kutter, E.F., Bostroem, J., Elger, C.E., Mormann, F., and Nieder, A. (2018). Single Neurons in the Human Brain Encode Numbers. Neuron *100*, 753-761.e4. 10.1016/j.neuron.2018.08.036.

33. Odabaee, M., Freeman, W.J., Colditz, P.B., Ramon, C., and Vanhatalo, S. (2013). Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes. NeuroImage *68*, 229–235. 10.1016/j.neuroimage.2012.11.062.

34. Stokes, M.G., Wolff, M.J., and Spaak, E. (2015). Decoding Rich Spatial Information with High Temporal Resolution. Trends Cogn. Sci. *19*, 636–638. 10.1016/j.tics.2015.08.016.

35. Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 10.3389/neuro.06.004.2008.

36. Nieder, A., Diester, I., and Tudusciuc, O. (2006). Temporal and Spatial Enumeration Processes in the Primate Parietal Cortex. Science *313*, 1431–1435. 10.1126/science.1130308.

37. King, J.-R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. *18*, 203–210. 10.1016/j.tics.2014.01.002.

38. Meck, W.H., and Church, R.M. (1983). A mode control model of counting and timing processes. J. Exp. Psychol. Anim. Behav. Process. *9*, 320–334. 10.1037/0097-7403.9.3.320.

39. Sun, C., Yang, W., Martin, J., and Tonegawa, S. (2020). Hippocampal neurons represent events as transferable units of experience. Nat. Neurosci. *23*, 651–663. 10.1038/s41593-020-0614-x.

40. Hannagan, T., Nieder, A., Viswanathan, P., and Dehaene, S. (2017). A random-matrix theory of the number sense. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *373*. 10.1098/rstb.2017.0253.

41. Lipton, J.S., and Spelke, E.S. (2004). Discrimination of Large and Small Numerosities by Human Infants. Infancy *5*, 271–290. 10.1207/s15327078in0503_2.

42. Xu, F., Spelke, E.S., and Goddard, S. (2005). Number sense in human infants. Dev. Sci. *8*, 88–101. https://doi.org/10.1111/j.1467-7687.2005.00395.x.

43. Cantlon, J.F. (2018). How Evolution Constrains Human Numerical Concepts. Child Dev. Perspect. *12*, 65–71. https://doi.org/10.1111/cdep.12264.

44. Wang, J. (Jenny), and Feigenson, L. (2021). Dynamic changes in numerical acuity in 4-month-old infants. Infancy *26*, 47–62. 10.1111/infa.12373.

45. de Hevia, M.D., Izard, V., Coubart, A., Spelke, E.S., and Streri, A. (2014). Representations of space, time, and number in neonates. Proc. Natl. Acad. Sci. *111*, 4809–4813. 10.1073/pnas.1323628111.

46. Lourenco, S.F., and Longo, M.R. (2010). General Magnitude Representation in Human Infants. Psychol. Sci. *21*, 873–881. 10.1177/0956797610370158.

47. de Hevia, M.D., and Spelke, E.S. (2010). Number-Space Mapping in Human Infants. Psychol. Sci. *21*, 653–660. 10.1177/0956797610366091.

48. Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., and Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. Dev. Sci. *21*, e12659. 10.1111/desc.12659.

49. Zorzi, M., and Testolin, A. (2018). An emergentist perspective on the origin of number sense. Philos. Trans. R. Soc. B Biol. Sci. *373*, 20170043. 10.1098/rstb.2017.0043.

50. Kim, G., Jang, J., Baek, S., Song, M., and Paik, S.-B. (2021). Visual number sense in untrained deep neural networks. Sci. Adv. *7*, eabd6127. 10.1126/sciadv.abd6127.

51. Giurfa, M. (2019). An Insect's Sense of Number. Trends Cogn. Sci. *23*, 720–722. 10.1016/j.tics.2019.06.010.

52. Potrich, D., Zanon, M., and Vallortigara, G. (2022). Archerfish number discrimination. eLife *11*, e74057. 10.7554/eLife.74057.

53. Kobylkov, D., Mayer, U., Zanon, M., and Vallortigara, G. (2022). Number neurons in the nidopallium of young domestic chicks. Proc. Natl. Acad. Sci. *119*, e2201039119. 10.1073/pnas.2201039119.

54. Messina, A., Potrich, D., Schiona, I., Sovrano, V.A., Fraser, S.E., Brennan, C.H., and Vallortigara, G. (2022). Neurons in the Dorso-Central Division of Zebrafish Pallium Respond to Change in Visual Numerosity. Cereb. Cortex *32*, 418–428. 10.1093/cercor/bhab218.

55. Jordan, K.E., Brannon, E.M., Logothetis, N.K., and Ghazanfar, A.A. (2005). Monkeys Match the Number of Voices They Hear to the Number of Faces They See. Curr. Biol. *15*, 1034–1038. 10.1016/j.cub.2005.04.056.

56. Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. Proc. Natl. Acad. Sci. *109*, 11860–11865. 10.1073/pnas.1204580109.

57. Ditz, H.M., and Nieder, A. (2015). Neurons selective to the number of visual items in the corvid songbird endbrain. Proc. Natl. Acad. Sci. *112*, 7827–7832. 10.1073/pnas.1504245112.

58. Merten, K., and Nieder, A. (2009). Compressed Scaling of Abstract Numerosity Representations in Adult Humans and Monkeys. J. Cogn. Neurosci. *21*, 333–346. 10.1162/jocn.2008.21032.

59. Jordan, K.E., and Brannon, E.M. (2006). Weber's Law influences numerical representations in rhesus macaques (Macaca mulatta). Anim. Cogn. *9*, 159–172. 10.1007/s10071-006-0017-8.

60. Piazza, M., Facoetti, A., Trussardi, A.N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., and Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. Cognition *116*, 33–41. 10.1016/j.cognition.2010.03.012.

61. Piazza, M., De Feo, V., Panzeri, S., and Dehaene, S. (2018). Learning to focus on number. Cognition *181*, 35–45. 10.1016/j.cognition.2018.07.011.

62. Piazza, M., Pica, P., Izard, V., Spelke, E.S., and Dehaene, S. (2013). Education Enhances the Acuity of the Nonverbal Approximate Number System. Psychol. Sci. *24*, 1037–1043. 10.1177/0956797612464057.

63. Georges, C., Guillaume, M., and Schiltz, C. (2020). A robust electrophysiological marker of spontaneous numerical discrimination. Sci. Rep. *10*, 18376. 10.1038/s41598-020-75307-y.

64. Fornaciai, M., and Park, J. (2018). Early Numerosity Encoding in Visual Cortex Is Not Sufficient for the Representation of Numerical Magnitude. J. Cogn. Neurosci. *30*, 1788–1802. 10.1162/jocn_a_01320.

65. Piazza, M., Mechelli, A., Price, C.J., and Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. Brain Res. *1106*, 177–188. 10.1016/j.brainres.2006.05.104.

66. Dormal, V., Andres, M., Dormal, G., and Pesenti, M. (2010). Mode-dependent and mode-independent representations of numerosity in the right intraparietal sulcus. NeuroImage *52*, 1677–1686. 10.1016/j.neuroimage.2010.04.254.

67. Cavdaroglu, S., Katz, C., and Knops, A. (2015). Dissociating estimation from comparison and response eliminates parietal involvement in sequential numerosity perception. NeuroImage *116*, 135–148. 10.1016/j.neuroimage.2015.04.019.

68. Cavdaroglu, S., and Knops, A. (2019). Evidence for a Posterior Parietal Cortex Contribution to Spatial but not Temporal Numerosity Perception. Cereb. Cortex *29*, 2965–2977. 10.1093/cercor/bhy163.

69. Arrighi, R., Togoli, I., and Burr, D.C. (2014). A generalized sense of number. Proc. R. Soc. B Biol. Sci. *281*, 20141791. 10.1098/rspb.2014.1791.

70. Burr, D.C., Anobile, G., and Arrighi, R. (2018). Psychophysical evidence for the number sense. Philos. Trans. R. Soc. B Biol. Sci. *373*, 20170045. 10.1098/rstb.2017.0045.

71. Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. Trends Cogn. Sci. *14*, 542–551. 10.1016/j.tics.2010.09.008.

72. Butterworth, B. (2018). The implications for education of an innate numerosity-processing mechanism. Philos. Trans. R. Soc. B Biol. Sci. *373*, 20170118. 10.1098/rstb.2017.0118.

73. Hyde, D.C., and Spelke, E.S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. Dev. Sci. *14*, 360–371. 10.1111/j.1467-7687.2010.00987.x.

74. vanMarle, K., and Wynn, K. (2009). Infants' auditory enumeration: Evidence for analog magnitudes in the small number range. Cognition *111*, 302–316. 10.1016/j.cognition.2009.01.011.

75. Brannon, E.M., Suanda, S., and Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. Dev. Sci. *10*, 770–777. https://doi.org/10.1111/j.1467-7687.2007.00635.x.

76. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods *51*, 195–203. 10.3758/s13428-018-01193-y.

77. Fló, A., Gennari, G., Benjamin, L., and Dehaene-Lambertz, G. (2022). Automated Pipeline for Infants Continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. Dev. Cogn. Neurosci. *54*, 101077. 10.1016/j.dcn.2022.101077.

78. Perrin, F., Pernier, J., Bertrand, O., and Echallier, J.F. (1989). Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. *72*, 184–187. 10.1016/0013-4694(89)90180-6.

79. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hämäläinen, M.S. (2014). MNE software for processing MEG and EEG data. NeuroImage *86*, 446–460. 10.1016/j.neuroimage.2013.10.027.

80. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al. (2013). MEG and EEG data analysis with MNE-Python. Front. Neurosci. *7*. 10.3389/fnins.2013.00267.

81. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

82. Grootswagers, T., Wardle, S.G., and Carlson, T.A. (2016). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. J. Cogn. Neurosci. *29*, 677–697. 10.1162/jocn_a_01068.

83. Gennari, G., Marti, S., Palu, M., Fló, A., and Dehaene-Lambertz, G. (2021). Orthogonal neural codes for speech in the infant brain. Proc. Natl. Acad. Sci. *118*. 10.1073/pnas.2020410118.

84. Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). LIBLINEAR: A Library for Large Linear Classification. J. Mach. Learn. Res. *9*, 1871–1874.

85. Al Roumi, F., Marti, S., Wang, L., Amalric, M., and Dehaene, S. (2021). Mental compression of spatial sequences in human working memory using numerical and geometrical primitives. Neuron *109*, 2627-2639.e4. 10.1016/j.neuron.2021.06.009.

86. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190. 10.1016/j.jneumeth.2007.03.024.